МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ,
МОЛОДЕЖИ И СПОРТА УКРАИНЫ

ХАРЬКОВСКАЯ НАЦИОНАЛЬНАЯ АКАДЕМИЯ ГОРОДСКОГО ХОЗЯЙСТВА

МАТЕМАТИЧЕСКИЕ МЕТОДЫ И МОДЕЛИ В ЭКОНОМИКЕ

Под общей редакцией В. И. Торкатюка

ХАРЬКОВ
ХНАГХ
2012

УДК 330.4:519.87:519.85
ББК 65в631+65.050.03
М33

Рецензенты:

А.П. Слесаренко, д-р физ.-мат. наук, проф. Института проблем машиностроения НАН Украины

М.В. Новожилова, д-р физ.-мат. наук, проф. Харьковского национального университета строительства и архитектуры

Рекомендовано к печати Ученым советом Харьковской национальной академии городского хозяйства, протокол № 5 от 27.01.2012 г.

М33 **Математические методы и модели в экономике**: монография/ В. И. Торкатюк, А. И. Колосов, В. Н. Бабаев и др.; под общ. ред. В. И. Торкатюка; Харьк. нац. акад. город. хоз-ва. − Х.: ХНАГХ, 2012. – 321 с.

ISBN 978-966-695-240-3

Рассматриваются основные методы финансовой математики: количественный анализ финансовых и кредитных операций, различные методы начисления процентов. Изложены и иллюстрируются примерами методы линейной и нелинейной оптимизации в экономике**.**

Представлена фрактальная гипотеза рынка, в соответствии с которой функционирование и стабильность рынков обеспечивается разнородностью инвесторов.

Для студентов экономических специальностей, аспирантов и научных работников, применяющих математические методы и модели в экономике.

УДК 330.4:519.87:519.85
ББК 65в631+65.050.03

ISBN 978-966-695-240-3

ВВЕДЕНИЕ

В современной экономической науке, обусловленной глобализацией мировой экономики и становлением общества нового типа – информационного общества − основными инструментами исследования и прогноза экономических явлений, становятся математические методы и модели. Использование в экономике математических методов связано с универсальностью последних.

Решение экстремальных экономических задач, направленных на достижение определенной цели, должно содержать следующие этапы: постановка проблемы, формирование целей и критериев, построение математических моделей, поиск оптимальных решений по различным целевым функциям, оценка полученного результата, корректировка модели.

Математическая модель (или формализованное описание проблемы) отражает связь между входными и выходными параметрами, на базе которой принимается управление процессом или объектом, удовлетворяющее заданной системе ограничений и доставляющее экстремальное значение целевой функции. Модель должна быть адекватной, т.е. соответствовать реальному процессу, отражать его основные закономерности, необходимые для принятия решения. Однако, сложная модель, учитывающая все особенности изучаемого процесса, может нарушить смысл моделирования, одна из целей которого − упростить постановку задачи, учитывая при этом ее существенные особенности.

Основы математического подхода к исследованию рыночной экономики были сформированы в XIX в. В XX в. учеными сделан существенный вклад в экономическую науку и математическое моделирование экономических процессов.

Одной из математических теорий, используемых в настоящее время для выработки решений на фондовых рынках, в которой рассматривается возможность математического моделирования риска с применением понятий теории вероятностей, является теория эффективных портфелей ценных бумаг.

Основные положения теории эффективных портфелей, направленной на решение практической задачи о рассредоточении капитала по различным видам ценных бумаг в условиях неопределенности, разработаны Г. Марковицем и изложены также в книге А. С. Шведова[32].

В основе метода построения множества эффективных портфелей лежит решение задач математического программирования.

В защиту вероятностных подходов следует отметить, что последствия экономического развития при строго детерминистском подходе не могут быть точно известны. Детерминистская постановка задач создает ощущение существования некоторого единственного оптимального плана. Однако этот план, полученный на основе случайно выбранных исходных данных, является случайным. Этот план не может устанавливаться однозначно, а должен формироваться как множество возможных планов, каждый из которых с некоторой вероятностью может оказаться оптимальным.

Одной из целей настоящего издания является ознакомление читателей с основами финансовой математики, то есть с математическим аппаратом, который используется для анализа финансовых операций.

Каждая финансовая, кредитная или коммерческая операция предполагает совокупность условий, согласованных ее участниками. К таким условиям относятся: сумма займа или инвестиций, цена товара, сроки, способы начисления процентов и погашения долга и т.д.

Совокупность влияющих на финансовую операцию факторов приводит к тому, что для оценивания ее конечного результата необходим специальный количественный анализ. Методы расчетов и составляют предмет финансовой математики. Количественный финансовый анализ имеет целью решение многих задач — от элементарного начисления процентов до анализа сложных кредитных и коммерческих операций в различных их постановках. В нем рассматриваются основные понятия, которыми оперируют в финансовых вычислениях, такие как ставка процента, современная или текущая стоимость платежа, методы наращения и дисконтирования платежей, идеи, лежащие в основе финансовых вычислений. Рассматриваются принцип финансовой эквивалентности платежей, количественный анализ при пересмотре условий финансовых соглашений, вопросы консолидации обязательств, инфляционное обесценение денег.

Издание содержит в основном традиционные методы финансово-экономических расчетов, основное внимание уделяется необходимому математическому инструментарию.

В настоящее время одной из основных становится задача создания системы оптимального планирования и управления на базе широкого применения математических методов в экономике как на уровне деятельности фирмы в условиях рынка, так и на уровне анализа экономической деятельности региона и страны.

Новейшие достижения математики и компьютерных технологий находят широкое применение в экономических исследованиях и планировании. Этому способствуют развитие таких разделов математики, как математическое программирова-ние, теория игр. Накоплен достаточный опыт постановки и решения экономических задач с помощью математических методов. Успешно развиваются методы оптимального планирования, составляющие сущность математического программирования.

Методы математического программирования − это методы отыскания и исследования экстремальных значений функций цели, на неизвестные величины которой наложены ограничения. В зависимости от вида рассматриваемых экономических задач на практике используются такие разделы как линейное, нелинейное, дискретное, динамическое программирование и др.

Для достижения цели имеются некоторые финансовые, трудовые и др. ресурсы, технологии и способы производства. Цель операции формализуется в виде критерия оптимальности и системы ограничений. Раздел математичес-кого программирования определяется как линейное программирование, если целевая функция и функции, входящие в систему ограничений, являются линейными относительно искомого плана. К методам линейной оптимизации относятся симплексный метод, модифицированный симплексный метод.

Задачи, в которых хотя бы одна из функций (целевая функция или какая-либо функция из системы ограничений) нелинейные, относятся к разделу нелинейного программирования: в частности, задачи выпуклого и квадратичного программирования, безусловной оптимизации, линейной аппроксимации. В экономике нелинейное программирование используется при расчете оптимальной партии выпуска деталей, управлении поставками и запасами, распределении ограниченных ресурсов, оптимизации показателей производственно-экономической деятельности и т.д.

Методом динамического программирования решаются задачи при условии, что параметры функции цели или системы ограничений изменяются во времени, или процесс принятия решений имеет многошаговый характер. Этим методом решаются задачи планирования, управления производством, поставками, распределения ограниченных ресурсов.

Методами дискретного программирования решаются задачи оптимизации с неделимостями, комбинаторного типа. Например, задачи выпуска неделимой продукции, календарного планирования, управления поставками при заданных транзитных нормах отпуска и т.д. Классические задачи целочисленной оптимизации: задача о назначениях, транспортная задача находят широкое применение на производстве, например, при закреплении машин за маршрутами, при назначении на должности с учетом запрещений для некоторых исполнителей.

Современные математические исследования в сочетании с компьютерными программами, производящими трудоемкие вычисления, позволяют использовать модели математического программирования в экономике.

Исследуются вопросы практического применения функций нескольких переменных в экономической теории и в задачах оптимизации. В частности, исследование функций многих переменных на условный экстремум и использование функции Лагранжа позволяют решать вопросы отыскания экстремума в задачах потребительского выбора со случайными эффективностями составных активов по критерию ожидаемой полезности.

Рассмотрены базовые понятия и основные законы экономической теории: функции полезности, спроса, потребления и предложения, производственная функция, законы убывающей доходности и убывающей полезности.

Приведенные в книге методы расчетов иллюстрируются разбором большого количества характерных примеров, актуальных задач и соответствующих экономических приложений.

Издание предназначено студентам экономических специальностей и будет полезно научным работникам, аспирантам, исследующим проблемы экономиики и применяющим математические методы и экономико-математическое моделирование.

Слава Тобі, Господи, що Ти створив усе потрібне простим, а все складне − непотрібним.

Григорій Сковорода

РАЗДЕЛ 1. НАРАЩЕНИЕ И ДИСКОНТИРОВАНИЕ ПО ПРОСТЫМ ПРОЦЕНТНЫМ СТАВКАМ

1.1. ПОНЯТИЕ ПРОЦЕНТА, ВИДЫ ПРОЦЕНТНЫХ СТАВОК

Основой количественного анализа каждой финансовой операции является финансовая математика. С одной стороны, в финансовом анализе процентная ставка является инструментом наращения суммы долга. С другой − это измеритель степени эффективности или доходности финансовой деятельности. Под процентными деньгами или, кратко, процентами в финансовых расчетах понимают [16, 17] абсолютную величину дохода от предоставления денег в долг в любой форме: в виде выдачи денежной ссуды, продажи в кредит, помещении денег на сберегательный счет, учет векселя, покупка сберегательного сертификата или облигаций и т.д. В какой бы форме не выступали проценты, это всегда конкретное проявление такой экономической категории, как ссудный процент.

В математике термин «процент» (от латинского pro centum — на сотню) означает сотую часть некоторого числа, или процент числа B есть $B/100$, то есть:

$$1\% \text{ числа } B = \frac{B}{100} = 0{,}01\ B.$$

Второе значение термина «процент» − экономическое.

Понятие процента широко используется во всех областях науки и техники, а для таких наук, как экономика, статистика и, особенно, бухгалтерский учет, значение этого математического инструмента трудно переоценить. В финансовой сфере слово «процент» представляет собой плату за использование денежных средств, предоставляемых одним лицом (кредитором) другому лицу (заемщику, дебитору, должнику), выраженную в сотых долях от суммы долга. Таким образом, экономическое значение процента более емкое, чем математическое.

Рассмотрим простейшие примеры.

Пример 1.

Температура воздуха утром составляла t_0=20°C, а днем t_1 =25°C. На сколько процентов она увеличилась?

Решение.

Температура воздуха увеличилась на величину равную $\Delta t = t_1 - t_0 = 5^{\circ} C$. Итак, надо определить, сколько процентов от числа t_0 составляет Δt.

Имеем $x = \frac{\Delta t}{t_0} \cdot 100 = 25$(%), то есть, температура увеличится на 25 % .

Пример 2.

В семейном бюджете, который составляет 4000 грн, затраты на питание составляют 25 % . Сколько денег расходуется на питание?

Составим пропорцию $\frac{x}{4000} = \frac{25}{100}$,

Отсюда $x = \frac{4000}{100} \cdot 25$ =1000(грн), т.е., на питание расходуется 1000 грн.

Пример 3.

20% жителей города (20000 чел.) составляют люди в возрасте от 40 до 55 лет. Сколько людей проживает в городе?

Решение.

Имеем пропорцию $\frac{20}{20000} = \frac{100}{x}$, откуда $x = \frac{20000}{20} \cdot 100$ =100000(чел).

Обобщая приведенные примеры, можно заметить, что существует три типа рассмотренных задач. Действительно, во всех случаях мы имели три числа, связанных пропорцией, одно из которых неизвестное.

Первая задача. Найти число B, если оно составляет C процентов от числа A. Находим

$$\frac{A}{100} = \frac{B}{C}, \text{ откуда } B = \frac{A}{100} \cdot C. \quad (1.1.1)$$

Вторая задача. Найти процент, который число B составляет от числа A. Имеем

$$\frac{B}{A}=\frac{C}{100}, \text{ откуда } C=\frac{B}{A}\cdot 100. \qquad (1.1.2)$$

Третья задача. Найти число A, если его C процентов равны числу B.

Тогда

$$\frac{C}{B}=\frac{100}{A}, \text{ откуда } A=\frac{B}{C}\cdot 100. \qquad (1.1.3)$$

Рассмотрим еще несколько примеров применения формул (1.1.1) − (1.1.3).

Пример 4.

Цену товара сначала повысили на 10 % , а потом уменьшили до исходной величины. На сколько процентов уменьшилась цена?

Решение.

Пусть C_0 − исходная цена. После повышения цены получим

$$C_1 = C_0 + C_0 \cdot 0{,}1 = C_0(1+0{,}1).$$

После этого цена уменьшилась на x%, тогда

$$C_2 = C_1 - C_1 \cdot x = C_1(1-x).$$

Из равенства $C_2 = C_0$ имеем $C_1(1-x) = C_0$, $C_0(1+0{,}1)(1-x) = C_0$,

$$1{,}1 - 1{,}1x = 1, \text{ отсюда } x = 0{,}0909 \text{ или } 9{,}1\%.$$

Неожиданный, на первый взгляд, результат связан, очевидно, с тем, что базой начисления скидки была не исходная цена C_0, а увеличенная цена $C_0 + 0{,}1C_0$.

Пример 5.

В декабре предприниматель получил прибыль C_0 = 1000 грн. Прибыль предпринимателя в январе увеличилась на 5 % , в феврале на 15%, а в марте уменьшилась на 10%. На сколько процентов выросла прибыль предпринимателя в I квартале?

Решение.

В январе прибыль достигала

C_1 = 1000 + 0,05$\cdot$ 1000 =1000(1 + 0,05)= 1050 (грн).

В феврале прибыль увеличилась еще на 15%, то есть составила

$$C_2 = C_1 + C_1 \cdot 0{,}15 = C_1(1+0{,}15) = 1207{,}5 (\text{грн}),$$

а в марте уменьшилась:

$$C_3 = C_2 - C_2 \cdot 0{,}1 = C_2(1-0{,}1) = 1086{,}75 \text{ (грн).}$$

Таким образом, увеличение составляет

$$C_3 - C_0 = 86{,}75(\text{ грн}) \text{ или } \frac{86{,}75}{1000} = 0{,}08675, \text{ то есть } 8{,}675\%.$$

Очевидно, что 5 + 15−10=10 $\neq$ 8,675, поскольку базами начисления процентов в феврале и марте был не декабрь, а, соответственно, январь и февраль.

Пример 6.

Цену товара сначала повысили на 10 %, а через день уменьшили на 10%. Какую цену будет иметь товар, если этот процесс будет длиться месяц?

Решение.

Имеем $C_1 = C_0 + 0{,}1C_0 = C_0(1+0{,}1),$ тогда

$$C_2 = C_1 - 0{,}1C_1 = C_0(1+0{,}1)(1-0{,}1) = C_0(1-0{,}01) = 0{,}99C_0.$$

Очевидно, что через месяц цена будет составлять

$$C_{30} = C_0(1-0{,}01)^{15} = 0{,}86C_0.$$

Если процесс будет длиться бесконечно $(n \to \infty)$, то получим

$$C_n = C_0(1-0{,}01)^{\frac{n}{2}} \underset{n\to\infty}{\longrightarrow} 0,$$

то есть, цена стремится к нулю, хотя ее повышение и понижение на каждом шаге происходит на одинаковую величину. Проценты за выданные в долг деньги — это один из важнейших элементов коммерческих, кредитных и инвестиционных контрактов, экономических и финансовых соглашений.

При заключении финансового или кредитного соглашения стороны (кредитор и заемщик) договариваются о размере процентной ставки.

Существуют различные способы начисления процентов (виды процентных ставок), зависящие от условий контрактов. Можно выделить ряд признаков, по которым различаются процентные ставки.

Проценты различаются по базе для их начисления. Применяется постоянная или последовательно изменяющаяся база для расчета. В последнем случае за базу принимается сумма, полученная на предыдущем этапе наращения, или дисконтирования, иначе говоря, проценты начисляются на проценты. При постоянной базе используют простые, при переменной — сложные процентные ставки. Проценты либо выплачиваются кредитору по мере их начисления, либо присоединяются к сумме долга. Процесс увеличения денег в связи с присоединением процентов к сумме долга называют *наращением* или ростом первоначальной суммы.

Важным также является выбор принципа расчетов процентов. Существует два таких принципа — наращение на сумму долга и скидка с конечной суммы задолженности. Соответственно применяют ставки наращения и учетные ставки. В специальной финансовой литературе проценты, полученные по ставке наращения, иногда называют *декурсивными*, по учетной ставке — *антисипативными*. Декурсивные проценты в большинстве случаев принято называть просто процентами.

Процентные ставки могут быть фиксированными (в контракте указывается их размер) или "плавающими" В последнем случае фиксируется не сама ставка, а изменяющаяся во времени база ("базовая ставка") и размер надбавки к ней — маржи. Размер маржи определяется рядом условий, в частности финансовым положением заемщика, сроком кредита и т.д. Размер маржи может быть постоянным на протяжении срока ссудной операции или переменным. В мировой практике, он обычно находится в пределах 0,5-5%

Интервал времени, к которому относится процентная ставка, называют *периодом начисления*. Ставка измеряется в процентах, в виде десятичной или натуральной дроби. В последнем случае она фиксируется в контрактах с точностью до 1/16 или даже 1/32.

Начисление процентов, как правило, производится дискретно, т.е. в отдельные (обычно равноотстоящие) моменты времени (дискретные проценты), причем, в качестве периодов начисления принимают год, полугодие, квартал, месяц. Иногда практикуют ежедневное начисление, а в ряде случаев удобно применять непрерывные проценты.

1.2. ВРЕМЕННАЯ ЗАВИСИМОСТЬ СТОИМОСТИ ДЕНЕГ

Прежде, чем рассмотреть основные виды финансовых операций и необходимый математический инструментарий, надо осознать экономическое содержание такого понятия, как процент, то есть, почему вообще, например, за кредит необходимо платить деньги (проценты).

Во-первых, отметим, что деньги непосредственно не удовлетворяют жизненных нужд и не могут потребляться. Итак, особа, имеющая деньги, должна сначала обменять их на товары потребления, услуги и т.п., то есть, воспользоваться потенциальной полезностью денег. Есть, однако, еще одна возможность — одолжить деньги (выдать кредит) другому лицу. Но тогда владелец лишается на определенный срок, возможности воспользоваться деньгами, ведь такая операция имеет для него смысл лишь тогда, когда он получит от должника плату (процент) за одолженные деньги. Ведь эта плата — компенсация за временную потерю владельцем потенциальной полезности денег.

В реальных случаях ситуация значительно усложняется такими факторами, как инфляционный риск и риск невыполнения обязательств должником. Инфляционный риск обусловлен процессами постепенного обесценения денег, может быть очень значительным, поскольку инфляция может достигать десятков и даже сотен процентов (гиперинфляция) на год. Итак, кредитор имеет право на дополнительную компенсацию — инфляционную премию. Невыполнение обязательств должником может произойти вследствие многих причин — банкротства, мошенничества, а также стихийного бедствия, боевых действий (форс-мажор). Очевидно, что и такой риск должен быть компенсированным.

Подытоживая сказанное, можно сделать такой вывод: настоящее владение деньгами имеет преимущество над владением деньгами в будущем. Итак, так называемое, положительное временное преимущество означает, что лучше получить деньги сегодня, чем спустя некоторое время в будущем.

В фундаментальной монографии Четыркина Е.М.[29] формулируется принцип неравноценности денег в разные моменты времени, а также приводится блестящий исторический пример, который иллюстрирует этот принцип (остров Манхеттен приблизительно 350 лет тому был куплен за 24 доллара, а сегодня стоимость только земли представляет свыше 40 млрд долларов). Таким образом, учет фактора времени или временной стоимости денег есть принципиально важным. Из принципа неравноценности денег непосредственно вытекает некорректность арифметических действий с денежными суммами, которые принадлежат к разным моментам времени, то есть некорректно, например, подытоживать прибыли за три месяца, чтобы вычислить прибыль за квартал.

В современном бухгалтерском учете такие „тонкости" игнорируются, однако анализ многих операций, особенно долгосрочных, нуждается в тщательном учете принципа неравноценности денег при вычислении, по крайней мере, самых важных параметров.

Дальше мы рассмотрим два метода решения этой проблемы — метод наращения и метод дисконтирования. В первом случае речь идет о вычислении будущей стоимости настоящей суммы денег. *Наращение* осуществляется путем добавления процентных денег к исходной сумме. *Дисконтирование* — обратная операция, которая состоит в вычислении современной стоимости будущей денежной суммы. Современную стоимость можно также рассматривать как сумму, которая при определенных условиях инвестирования возрастет в будущем до заданной величины.

1.3. ФОРМУЛА ПРОСТЫХ ПРОЦЕНТОВ

Рассмотрим операцию — размещение денег на банковском депозите на определенный срок. Пусть P — исходная или настоящая сумма, которая размещена на депозит, S — наращенная сумма, которую будет уплачено в конце операции. Под наращенной суммой ссуды (долга, депозита, других видов инвестированных средств) понимается первоначальная ее сумма вместе с начисленными на нее процентами к концу срока. Разность

$$S - P = I \qquad (1.3.1)$$

— абсолютный прирост денежной суммы, будем называть общей суммой процентных денег.

Для количественной оценки временной стоимости денег используют два параметра:

$$i = \frac{I}{P} = \frac{S-P}{P} - \textit{процентная ставка}\ , \qquad (1.3.2)$$

$$d = \frac{I}{S} = \frac{S-P}{S} - \textit{дисконтная ставка}. \qquad (1.3.3)$$

Данные параметры характеризуют отношение абсолютного прироста I к соответствующей базовой величине — настоящей сумме P или наращенной сумме S. Таким образом, через определенный период банком будет уплачена сумма, которая отличается от настоящей.

Используя формулу, (1.3.1), находим:

$$S = P + I = P + iP = P(1+i). \qquad (1.3.4)$$

Очевидно, что для любого $i > 0$ выполняется неравенство $S > P$.

В конкретных финансовых расчетах используются различные методы начисления процентов. Если базовая величина (база начисления) есть постоянная, то проценты называют простыми. Так, при размещении денег, еще на один срок формула (1.3.4) дает: $S = (P + iP) + iP = P + 2iP$.

1.4. НАРАЩЕНИЕ И ДИСКОНТИРОВАНИЕ С ИСПОЛЬЗОВАНИЕМ ПРОСТОЙ ПРОЦЕНТНОЙ СТАВКИ

Наиболее распространенной схемой, используемой в краткосрочных операциях, является наращение с использованием простых процентов. Наращенная сумма определяется умножением первоначальной суммы долга на множитель наращения, который показывает, во сколько раз наращенная сумма больше первоначальной. Расчетная формула зависит от вида применяемой процентной ставки и условий наращения. При наращении простых процентов по ставке i (каждая следующая сумма больше предыдущей на долю i от начальной суммы P, т.е. на iP). К концу единичного промежутка начис-ления сумма P возрастет на iP и станет $P_1 = P + iP = P(1+i)$, к концу 2-го промежутка начисления эта сумма возрастет еще на iP и станет $P_2 = P_1 + iP = P(1+i) + iP = P(1+2i)$, и т.д. К концу n-го промежутка начисления наращенная сумма станет $P_n = P(1+ni)$. Таким образом, последовательность наращенных сумм $P, P_1, \ldots, P_n$ есть арифметическая прогрессия с начальным членом P и разностью iP. Формула наращения по простым процентам или, кратко, формула простых процентов имеет вид:

$$S = P(1+ni), \tag{1.4.1}$$

где S – наращенная сумма или сумма в конце срока, P – первоначальная сумма (ссуда), n – срок начисления процентов (срок ссуды), i – ставка наращения или ставка процентов за единицу времени (десятичная дробь). Срок ссуды чаще всего измеряется в годах, соответственно i - годовая ставка. Начисленные за один год проценты составят величину Pi. За n лет будут начислены проценты в сумме $I = Pni$. Множитель $M_n^{(1)} = 1 + ni = \frac{S}{P}$ в формуле (1.4.1) называется множителем наращения простых процентов. Разность наращенной суммы и начальной называется процентными деньгами. При наращении простых процентов процентные деньги растут в арифметической прогрессии. Графически это показано на рис. 1.1, где P − начальная сумма, отрезки P_kT_k — наращенные суммы и отрезки P_kM_k — процентные деньги.

Пример 1.

Годовая ставка простых процентов равна 12%. Через сколько лет начальная сумма удвоится?

Решение: Требуется решить неравенство: $1+0{,}12n \geq 2$, т.е. $0{,}12n \geq 1$. Получаем $n \geq 1/0{,}12$. Ответ: через 9 лет.

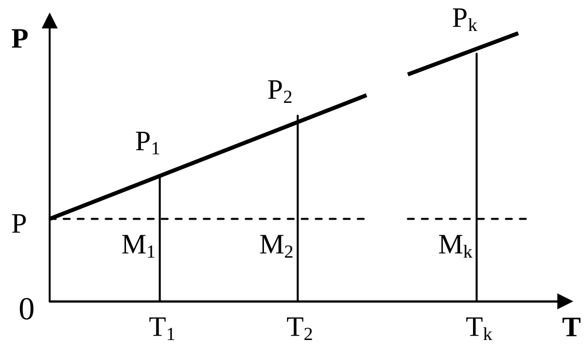


Рис. 1.1

Увеличение процентной ставки или срока в k раз приводит к увеличению множителя наращения в $(1+kni)/(1+ni)$ раз.

Пример 2.

Вычислить проценты и сумму накопленного долга, если ссуда равна 7000 грн., срок — четыре года, проценты простые, ставка — 10% годовых. Находим $I = 7000 \cdot 4 \cdot 0{,}1 = 2800$ грн.; S = 7000 + 2800 = 9800 грн.

Увеличим ставку в два раза. Сумма процентов при этом удвоится, однако наращенная сумма увеличится в $(1 + 2 \cdot 4 \cdot 0{,}1) / (1 + 4 \cdot 0{,}1) = 1{,}286$ раза.

Заметим, что временной базой для начисления процентов в большинстве случаев является дискретный интервал времени (год, квартал или месяц). В современных условиях, например, в банковской сфере, осуществляются операции продолжительностью в несколько дней и даже часов, т.е. проценты начисляются почти непрерывно. Естественно, что в первом случае проценты называют дискретными, а во втором — непрерывными

Обычно к наращению по простым процентам прибегают при выдаче краткосрочных ссуд (на срок до одного года) или в случаях, когда проценты не присоединяются к сумме долга, а периодически выплачиваются кредитору. Поскольку ставка фиксируется в контракте в расчете за год, то при сроке ссуды менее года необходимо определить, какая часть годового процента уплачивается кредитору. Сходная проблема возникает и в других случаях, когда срок ссуды меньше периода начисления.

Формула наращения простых процентов $S = P(1+ni)$, выведенная для целых положительных n, может применяться и для промежутков времени, выраженных в нецелых числах через базовый период. Сумма S, наращенная по ставке i простых процентов, через промежуток t начисления станет

$$S = P(1+\frac{t}{K}i), \tag{1.4.2}$$

где t − количество дней ссуды, K − продолжительность года в днях или временная база, $n = \frac{t}{K}$. Наращенная сумма S является линейной функцией времени.

При расчете простых процентов предполагают, что K = 360 (12 месяцев по 30 дней) или K = 365, 366 дней. Если K = 360, то получают обыкновен-ные, или коммерческие, проценты; при использовании действительной продолжительности года (365, 366) получают точные проценты.

Число дней ссуды также можно измерить приближенно или точно. В первом случае продолжительность ссуды определяется из условия, согласно которому любой месяц принимается равным 30 дням. Точное число дней ссуды определяется путем подсчета числа дней между датой выдачи ссуды и датой ее погашения. День выдачи и день погашения считаются за один день. Для подсчета числа дней можно воспользоваться табл. Приложение.

На практике применяются три варианта расчета простых процентов:

а) *точные проценты с точным числом дней ссуды*. Этот вариант, естественно, дает самые точные результаты. Данный способ применяется центральными банками многих стран и крупными коммерческими банками. Обычно он обозначается как 365/365 или АСТ/АСТ;

б) *обыкновенные проценты с точным числом дней ссуды*. Этот метод, иногда называемый банковским, распространен в ссудных операциях коммерческих банков. Он обозначается как 365/360 или АСТ/360. Этот вариант дает несколько больший результат, чем применение точных процентов. Заметим, что при числе дней ссуды, превышающем 360, данный способ приводит к тому, что сумма начисленных процентов будет больше, чем предусматри-вается годовой ставкой. Например, если $t = 364$, то $n = 364/360 = 1{,}011$;

в) *обыкновенные проценты с приближенным числом дней ссуды.* Такой метод применяется тогда, когда не требуется большой точности, например при промежуточных расчетах. Этот метод обозначается как 360/360.

Вариант расчета с точными процентами и приближенным числом дней ссуды лишен смысла и не применяется.

Заметим, что поскольку точное число дней ссуды в большинстве случаев, однако, не всегда больше приближенного (в чем легко убедиться, определив среднее число дней в месяце, которое равно 30,58), то проценты с точным числом дней обычно дают больший рост.

Математически корректным является, безусловно, метод с использованием точного количества дней в году (365/365) и точным количеством дней ссуды (точные проценты с точным числом дней ссуды).

Пример 3.

Заем в размере 2000 грн. выдан на 45 дней под 12 % годовых. Вычислить сумму, которую должен возвратить должник в конце срока.

Согласно формуле (1.4.2) имеем

$$S = 2000(1 + \frac{45}{365} \cdot 0{,}12) = 2029{,}59(\text{ грн}).$$

Пример 4.

Сумму в 3000 грн. размещено на депозитном счете с начислением 18% годовых. Через какое время исходная сумма возрастет до 4 000 грн.?

Решение: Согласно (1.4.2) имеем

$$t = \frac{S - P}{P} \cdot \frac{K}{i} = \frac{4000 - 3000}{3000} \cdot \frac{365}{0{,}18} = 676\,(\text{дней}).$$

Пример 5.

Какую процентную ставку необходимо использовать, чтобы сумма в 500 грн, размещенная на депозитном счете на 155 дней, превратилась в 600 грн?

$$i=\frac{S-P}{P}\cdot\frac{K}{t}=\frac{600-500}{500}.\frac{365}{155}=0,47 \text{ или 47\% годовых.}$$

Пусть срок ссудной операции находится в смежных календарных периодах. Необходимость деления общей суммы процентов между периодами возникает в финансовом анализе деятельности предприятия, в бухгалтерском учете, при налогообложении. Если общий срок ссуды захватывает два смежных календарных периода, причем на первый период приходится срок n_1, на второй — n_2, то

$$I=I_1+I_2=P_{n_1}i+P_{n_2}i.$$

В кредитных соглашениях иногда предусматриваются изменяющие-ся во времени процентные ставки. В случае простых ставок наращенная на конец срока сумма определяется формулой

$$S=P(1+n_1i_1+n_2i_2+...+n_mi_m)=P(1+\sum_{k=1}^{m}n_ki_k) \qquad (1.4.3)$$

где i_k — ставка простых процентов в периоде k, $k = 1,2,...,m$;

n_k — продолжительность периода; $n=\sum_{k=1}^{m}n_k$.

Формула (1.4.3) — обобщение (1.4.1) на случай, когда период начисления состоит из m слагаемых, которым соответствуют различные процентные ставки $i_1,i_2,...,i_m$.

Очевидно, что за период n_k сумма процентных денег составит:

$$I_k=Pn_ki_k.$$

За весь период n сумма процентных денег равна: $I=P\sum_{k=1}^{m}n_ki_k$.

Итак, $$S=P+I=P(1+\sum_{k=1}^{m}n_ki_k). \qquad (1.4.4)$$

Соответствующий множитель наращения имеет вид $M_n^{(1)}=\frac{S}{P}=1+\sum_{k=1}^{m}n_ki_k$.

Пример 6.

Заем в сумме 3000 грн выдано на 1 год. В первом квартале исполь-зуется ставка 10%, во втором — 15%, а в третьем и четвертом — 20% годовых. Найти сумму, которую необходимо возвратить.

Решение: Согласно (1.4.4) имеем

$$S = 3000(1+\frac{1}{4}\cdot 0{,}1+\frac{1}{4}\cdot 0{,}15+\frac{2}{4}\cdot 0{,}2) = 3487{,}5\,(\text{грн}).$$

Рассмотрим теперь обратную задачу — вычисление настоящей величины денежной суммы P, если известна наращенная сумма S.

Такая задача возникает, например, когда необходимо вычислить сумму для размещения на депозитном счете, чтобы через определенный срок она достигла необходимой величины. В общем случае необходимо, исходя из величины денежной суммы в будущем S, вычислить величину денежной суммы P в любой предыдущий момент.

Такая задача решается с помощью метода *дисконтирования* или *приведения*. Одна из возможностей состоит в использовании равенства (1.4.1), разрешив которое относительно переменной P получим

$$P = \frac{S}{1+ni} = SD_n^{(1)}. \qquad (1.4.5)$$

Метод, состоящий в определении настоящей суммы по наращенной сумме, называется *методом математического дисконтирования.*

Коэффициент $D_n^{(1)}$ называется дисконтным множителем и, очевидно,

$$D_n^{(1)} = \frac{1}{1+ni} < 1$$

для любых n и i; $D_n^{(1)} = \dfrac{1}{1+ni} < 1$, следовательно, настоящая сумма P всегда меньше наращенной суммы S. Дисконтный множитель показывает какую долю составляет первоначальная сумма ссуды в окончательной величине долга. Дисконт суммы S равен: $D = S - P$.

Зависимость (1.4.5) не является линейной.

Пример 7.

На момент окончания кредитного соглашения должник уплатил 10000 грн. Процентная ставка — 10 % годовых. Какую сумму кредита получено, если продолжительность соглашения представляет: а) 90 дней, б) 270 дней?

Решение: В соответствии с формулой (1.11) имеем

а) $P = \dfrac{10000}{1+\dfrac{90}{365}\cdot 0{,}1} = 9759{,}90$ (грн).

б) $P = \dfrac{10000}{1+\dfrac{270}{365}\cdot 0{,}1} = 9311{,}84$ (грн).

Полученные результаты подтверждают тот факт, что настоящая стоимость денежной суммы существенно зависит от выбора момента, к которому осуществляется приведение.

Пример 8.

Для погашения кредита в сумме 2500 грн необходимо уплатить 5000 грн. Вычислить продолжительность кредитного соглашения, если использовалась процентная ставка 10% годовых.

Решение: Согласно формуле (1.5.1) имеем

$$t = \frac{S-P}{Pi}\cdot K = \frac{5000-2500}{2500}\cdot\frac{365}{0{,}1} = 3650 \text{ (дней)=10(лет)},$$

то есть, при ставке 10% настоящая величина суммы S равняется лишь половине ее будущей величины, если продолжительность кредитного соглашения составляет10 лет.

Пример 9.

Руководитель предприятия имеет целью модернизацию производства. Фирма-поставщик предлагает два варианта приобретения необходимого оборудования, а именно: предварительную оплату в сумме 80000 грн или 20000 грн в начале операции, а еще через полгода остаток — 65000 грн.

Необходимо сравнить эти предложения, если процентная ставка i = 12 % годовых.

Решение:

Для сравнения двух сумм 60000 и 65000 необходимо воспользоваться формулой (1.4.5). Имеем:

$$P = \frac{65000}{1+\dfrac{1}{2}\cdot 0{,}12} = 61320{,}75 \text{ (грн).}$$

Принимая первое предложение, предприятие сэкономит

$$61320{,}75 - 60000 = 1320{,}75 \text{ (грн).}$$

1.5. ДИСКОНТИРОВАНИЕ И НАРАЩЕНИЕ С ИСПОЛЬЗОВАНИЕМ ПРОСТОЙ ДИСКОНТНОЙ СТАВКИ

Практическое значение имеет задача обратная наращению процентов когда по заданной сумме S, соответствующей концу финансовой операции, требуется найти исходную сумму P. Расчет P по S называется *дисконтированием* суммы S. Величину P, найденную дисконтированием, называют *современной величиной* (*текущей стоимостью*) суммы S. Проценты в виде разности $D = S - P$ называются *дисконтом* или *скидкой*. Процесс начисления и удержания процентов вперед (в виде дисконта) называют *учетом*. Таким образом, в практике используются два принципа расчета процентов: (1) путем наращения суммы ссуды и (2) устанавливая скидку с конечной суммы долга.

В большинстве случаев фактор времени учитывается в финансовых контрактах именно с помощью дисконтирования. Величина P *эквивалентна* сумме S в том смысле, что через определенный период времени и при заданной ставке процентов она в результате наращения станет равной S. Поэтому операцию дисконтирования называют также приведением. Но понятие приведения шире, чем дисконтирование.

Приведение − это определение любой стоимостной величины в некоторый момент времени. Если некоторая сумма приводится к более ранней дате, чем текущая, то применяется дисконтирование, если же речь идет о более поздней дате, то − наращение. Известны два вида дисконтирования: математическое дисконтирование и банковский (коммерческий) учет.

1.5.1. Математическое дисконтирование

Этот вид дисконтирования представляет собой решение задачи, обратной наращению первоначальной ссуды. Если в прямой задаче $S = P(1+ni)$, то в обратной $P = S\dfrac{1}{1+ni}$.

1.5.2. Банковский или коммерческий учет

Операция учета (учета векселей) заключается в том, что банк до наступления срока платежа по векселю или другому платежному обязательству покупает его у владельца (являющегося кредитором) по цене ниже той суммы, которая должна быть выплачена по нему в конце срока, т.е. приобретает (учитывает) его с дисконтом. Для расчета процентов при учете векселей применяется *учетная* ставка, которую мы обозначим символом d. По определению, простая годовая учетная ставка находится как

$$d=\frac{S-P}{S}, \quad \text{откуда} \quad P=S(1-d).$$

В общем случае:

$$P=S(1-nd), \text{ или } P=S(1-\frac{t}{K}d). \qquad (1.5.1).$$

Итак, вместо процентной ставки i тут используется учетная (дисконтная) ставка d. Размер дисконта или учета, удерживаемого банком, равен: $D=Snd$.

1.5.3. Удержание простых процентов

Пусть необходимо взять в банке кредит в размере 10000грн. Банк предлагает следующие условия: процентная ставка 10% годовых, проценты в размере 1000грн. будут удержаны сейчас, а через год нужно будет вернуть все 10000грн. Эта операция называется *удержанием процентов.* В данной финансовой операции все в пользу банка. Во-первых, проценты уже удержаны. Во-вторых, доходность этой операции для банка больше, чем объявленные 10%.

Действительно, доходность операции для банка равна 1000/9000~11,11%. Поэтому подобную операцию – удержание процентов с конечной суммы – кредиторы применяют довольно часто.

Долговая расписка, содержащая обязательство выплатить определенную денежную сумму (номинал векселя) в конкретный срок, называется *векселем*. *Учет векселя* – это оплата векселя с дисконтом, т.е. со скидкой с его номинала.

Рассмотрим удержание простых процентов. Пусть ставка удержания – d (доля), тогда за каждый год удерживается одна и та же величина – доля d с конечной суммы S, так что если кредит выдается на n лет, то будет удержано ndS и оставшаяся после удержания сумма: $P=S(1-nd)$. Оставшиеся после удержания суммы образуют убывающую арифметическую прогрессию.

Пример 1.

С суммы 1000 удерживаются проценты по ставке 5%. Выписать оставшиеся суммы.

Промежутки удержания	−5	−4	−3	−2	−1	0
Простые проценты	750	800	850	900	950	1000
	(арифметическая прогрессия)					

Величина $D_n^{(2)} = \frac{P}{S} = 1 - nd$ называется *дисконтным множителем* для простых процентов.

Срок n измеряет период времени от момента учета векселя до даты его погашения в годах. Дисконтирование по учетной ставке производится чаще всего при условии, что год равен 360 дням.

С математической точки зрения оба дисконтных $D_n^{(1)}, D_n^{(2)}$ множителя выполняют ту же самую функцию — вычисляют отношение P/S.

Однако с финансовой точки зрения отличие является существенным. Напомним, что проценты, начисленные с помощью процентной ставки (декурсивные проценты), платятся после завершения периода начисления, исходя из величины суммы P.

1.5.4. Наращение по учетной ставке

Учетная ставка может использоваться для наращения, т.е. для расчета S по P. В этом случае из формулы (1.5.1) следует, что

$$S = P\frac{1}{1-nd} = PM_n^{(2)},$$

где множитель наращения по учетной ставке d равен $M_n^{(2)} = \frac{1}{1-nd}$.

Пример 2.

Владелец продал банку вексель на сумму 2000 грн за 36 дней до погашения. Найти сумму, которую получил владелец и прибыль банка (дисконт), если использовалась дисконтная ставка d =20 % .

Решение:

Согласно формуле (1.5.1) имеем $P = 2000(1 - 0,2 \cdot \frac{36}{360}) = 1960$ (грн),

откуда $D = S - P =$ 2000−1960 = 40 (грн).

Рассмотрим теперь, каким образом начисляются проценты с помощью дисконтной ставки (антисипативные проценты). Типичная операция состоит в приобретении и следующей реализации некоторого финансового обязательства. Например, вексель можно приобрести до истечения срока уплаты за цену, которая меньше номинальной, и получить вознаграждение (дисконт) после реализации векселя в конце срока. Итак, процентные деньги (дисконт) начисляются исходя из наращенной суммы S. Следует также обратить внимание на то, что множитель наращения $M_n^{(2)}$ и формула (1.5.1) имеют смысл лишь при условии $1-nd>0$.

1.5.5. Определение продолжительности ссуды

Пусть требуется найти временной интервал, за который исходная сумма при заданной ставке процентов вырастет до нужной величины, или срок, обеспечивающий определенный дисконт с заданной величины. Другими словами, речь идет об определении n из равенств (1.4.5) и (1.5.1).

При использовании простой ставки наращения i из (1.4.5) получаем

$$n=\frac{S-P}{Pi}, \tag{1.5.2}$$

а при учетной ставке d из (1.12) имеем

$$n=\frac{S-P}{Sd}. \tag{1.5.3}$$

Формулы (1.5.2) и (1.5.3) дают срок, измеряемый в годах, но простые ставки в основном используются в краткосрочных операциях, когда срок исчисляется днями. В этом случае срок финансовой операции в днях выражается как $t=nK$, где K − временная база.

1.5.6. Определение процентной и учетной ставок

Процентная ставка может служить мерой доходности операции, критерием выбора наиболее выгодных условий. Из тех же формул (1.4.5) и (1.5.1) получаем ставку наращения i и учетную ставку d соответственно.

$$i=\frac{S-P}{Pn}=\frac{S-P}{Pt}K, \tag{1.5.4}$$

$$d=\frac{S-P}{Sn}=\frac{S-P}{St}K. \tag{1.5.5}$$

Напомним, что срок n в двух последних формулах имеет различный смысл: в первом случае это весь срок операции, а во втором − оставшийся срок до погашения.

Пример 3

Определить доходность операции для кредитора, если им предоставлена ссуда в размере 2 млн. грн. на 200 дней и контракт предусматривает сумму погашения долга 2,5 млн. грн. Доходность выразить в виде простой ставки процентов i и учетной ставки d. Временную базу принять равной K=360 дней.

Решение.

$$i = \frac{S-P}{Pt}K = \frac{2,5-2}{2\cdot 200}\cdot 360 = 0,45,\text{ т.е. }45\%.$$

$$d = \frac{S-P}{St}K = \frac{2,5-2}{2,5\cdot 200}360 = 0,36\text{ ,т.е.}36\%.$$

Пример 4

Владелец векселя на сумму 3000 грн продал его банку за 4 года до срока погашения. Банк выплатил владельцу 2500 грн. Найти дисконтную ставку, которая использовалась, а также стоимость векселя за 2 года до погашения.

Решение:

Согласно формуле (1.5.3) имеем:

$$d = \frac{S-P}{Sn} = \frac{3000-2500}{3000\cdot 4} = 0,0417,\text{ или }4,17\%.$$

Стоимость векселя за 2 года до погашения равна

$$P = S(1-nd) = 3000(1-2\cdot 0,0417) = 2749,80\text{ (грн).}$$

Иногда размер дисконта в контрактах фиксируется за весь срок ссуды в виде доли (или процента) от суммы погасительного платежа.

Таким образом, значения процентной и дисконтной ставок задаются в неявном виде. Нетрудно вывести формулы, с помощью которых значения этих ставок можно вычислить.

Пусть $S-$ размер погасительного платежа, d_n- доля этого платежа, определяющая величину дисконта за весь срок ссуды *n*. Определить годовые ставки *i* и d, которым эквивалентны данные условия.

Итак, S − сумма возврата в конце срока ссуды, $P=S(1-d_n)$ выдаваемая ссуда в момент заключения договора.

$$i = \frac{S-P}{Pn} = \frac{S-S(1-d_n)}{S(1-d_n)n} = \frac{d_n}{(1-d_n)n},\ d = \frac{S-P}{Sn} = \frac{S-S(1-d_n)}{Sn} = \frac{d_n}{n}.$$

Пример 5.

Кредитор и заемщик договорились, что из суммы кредита, выданного на 200 дней, сразу удерживается дисконт в размере 25% указанной суммы. Требуется определить цену кредита в виде простой годовой учетной ставки *d* и годовой ставки простых процентов *i*. Считать временную базу *K* равной 365 дням.

Решение.

$$d == \frac{d_n}{n} = \frac{0,25}{200/365} = 0,4563, \text{ или } 45,63\%,$$

$$i = \frac{d_n}{(1-d_n)n} = \frac{0,25 \cdot 365}{(1-0,25) \cdot 200} = 0,61, \text{ или } 61\%.$$

1.5.7. Сравнение ставки наращения и учетной ставки

Операции наращения и дисконтирования по своей сути противоположны, но ставка наращения и учетная ставка могут использоваться для решения обеих задач. В этом случае, в зависимости от применяемой ставки, нужно различать прямую и обратную задачи.

Прямая и обратная задачи

Ставка	Прямая задача	Обратная задача
наращения i	наращение $S = P(1+ni)$	дисконтирование $P = S/(1+ni)$
учетная d	дисконтирование $P = S(1-nd)$	наращение $S = P/(1-nd)$

1.5.8. Совмещение начисления процентов по ставке наращения и дисконтирования по учетной ставке

В том случае, когда учету подлежит долговое обязательство, предусматривающее начисление простых процентов на первоначальную сумму долга, необходимо решить две задачи: (1) определить конечную сумму долга на момент его погашения; (2) вычислить сумму, получаемую при учёте, путем дисконтирования конечной суммы долга, применяя учетную ставку, действующую в момент учета.

Решение двух этих задач можно записать в виде одной формулы, содержащей наращение по ставке простых процентов, фигурирующей в долговом обязательстве, и дисконтирование по учетной ставке:

$$S = P(1+n_1 i)(1-n_2 d), \tag{1.5.6}$$

где $P-$ первоначальная сумма ссуды, $S-$сумма, получаемая при учете обязательства, n_1-общий срок платежного обязательства, в течение которого начисляются проценты, n_2- срок от момента учета до погашения долга $n_2 < n_1$.

Пример 6.

Обязательство в сумме 20000 грн необходимо возвратить с процентами (12%) через 3 года. Банк согласился с досрочным (через 2 года) погашением долга, если будет использована дисконтная ставка d (18%).Какая сумма будет возвращена банком?

В соответствии с формулой (1.5.6) получим

S=20000·(1+3·0,12)·(1−(3−2)·0,18)=20000·1,36·0,82=2304(грн).

Таким образом, собственник обязательства получил не

20000·(1+2·0,12)=24800(грн),

а сумму, меньшую на величину 24800 −22304=2496(грн).

1.5.9. Реинвестирование по простым ставка.

При инвестировании средств в краткосрочные депозиты иногда прибегают к последовательному повторению наращения по простым процентам в пределах заданного общего срока, т.е. к реинвестированию полученных на каждом этапе наращения средств. Наращенная сумма для всего срока будет находиться по формуле

$$S = P(1+n_1 i_1)(1+n_2 i_2)...(1+n_k i_k) \qquad (1.5.7)$$

где i_k — ставки, по которым производится реинвестирование на k-м периоде начисления.

Если периоды начисления и ставки не изменяются во времени, то вместо формулы (1.5.7) получим:

$$S = P(1+ni)^m,$$

где m — количество повторений реинвестирования.

Пример 7.

20000 грн. инвестированы 1 ноября на месячный депозит под 20% годовых. Какова наращенная сумма, если операция повторяется два раза? В случае начисления точных процентов получим

$$S = 20000 \cdot \left(1+\frac{30}{365}\cdot 0,2\right)\cdot\left(1+\frac{31}{365}\cdot 0,2\right) = 20674\ грн.$$

Начисление обыкновенных процентов дает

$$S = 20000 \cdot \left(1+\frac{30}{365}\cdot 0,2\right)^2 = 20672\ грн.$$

Когда боги хотят наказать нас,
они внемлют нашим молитвам.
Уайльд.

1.6. ФИНАНСОВАЯ ПИРАМИДА

Многие из нас не могут справиться с искушением быстро заработать. И когда возникает возможность «заработать» за месяц прибыль в размере 20 % и даже 30 %, более того, и знакомые уже получали такие проценты, то стоит ли сомневаться?

Действительно, так называемая «пирамида» некоторое время платит «огромные» проценты, однако ее существование всегда имеет трагический финал для вкладчиков.

Время существования пирамиды зависит от многих параметров, а именно: величины процентной ставки, минимального срока депозита, количества вкладчиков и т.п.

Рассмотрим простейший вариант, когда сумма ежемесячных поступлений стала P, ежемесячно вкладчики получают i простых процентов, но основную сумму, которая размещена на депозите, можно забрать лишь по истечении N месяцев [9].

Пусть P_n — сумма, накопленная в кассе «пирамиды» в течение n месяцев. Очевидно, что по истечении первого месяца в кассе появится сумма

$P_1 = P$,

второго — $P_2 = P + P(1-i)$,

третьего — $P_3 = P + P(1-i) + P(1-2i)$,

четвертого — $P_4 = P + P(1-i) + P(1-2i) + P(1-3i)$.

По истечении n месяцев $(n < N)$ в кассе будет сумма

$$P_n = P + P(1-i) + P(1-2i) + \ldots + P(1-(n-1)i) =$$
$$= P(n - i(1 + 2 + \ldots + (n-1))).$$

Используя формулу для суммы членов арифметической прогрессии во внутренних скобках, находим

$$P_n = P(n - \frac{in}{2}(n-1)).$$

Для определения максимального значення P_n достаточно найти производную и приравнять ее нулю. Имеем

$$P_n' = P\left(n - \frac{i}{2}n^2 + \frac{i}{2}n\right)' = 0,$$

Или $1 - in + \frac{i}{2} = 0$, откуда:

$$n_0 = \frac{1}{2} + \frac{1}{i}. \tag{1.6.1}$$

Следовательно, сумма P_n будет возрастать до максимального значения

$$P_{n_0} = P_{\max} = \frac{P}{2}\left(1 + \frac{1}{i} + \frac{i}{4}\right). \tag{1.6.2}$$

Деятельность пирамиды целесообразна до тех пор, пока производная накопленной суммы $P_n' \geq 0,$ откуда:

$$n \leq \frac{1}{2} + \frac{1}{i}. \tag{1.6.3}$$

Как следует из (1.6.3), продолжительность существования «пирамиды» тем меньше, чем больше процентная ставка по депозиту. Минимальная продолжительность существования «пирамиды» − 1 месяц, отсюда

$$\frac{1}{2} + \frac{1}{i} = 1, \text{ то есть } i_{\max} = 50\%.$$

Рассмотрим конкретный случай.

Пусть организаторы платят вкладчикам 20 % ежемесячно, минимальный срок депозита будет, очевидно, больше n_0, например, $N = n_0 + 1,$ то есть

$$N = \left(\frac{1}{2} + \frac{1}{0,2}\right) + 1 = 5,5 + 1 = 6,5 \approx 7 \text{ (месяцев)}.$$

Согласно формулам (1.6.2) и (1.6.3) через 5,5 месяца организаторы «пирамиды» накопят сумму $P_{\max} = \frac{P}{2}\left(1 + \frac{1}{0,2} + \frac{0,2}{4}\right) = 3,025P$. После этого организаторы исчезают вместе с $P_{\max}$, поскольку сумма P начнет уменьшаться и платить проценты станет нечем.

1.7. КОНВЕРСИЯ ВАЛЮТЫ И НАЧИСЛЕНИЕ ПРОЦЕНТОВ

Рассмотрим совмещение конверсии (обмена) валюты и наращение простых процентов, сравним результаты при непосредственном размещении денежных средств в депозиты или после предварительного обмена на другую валюту [11].

При этом имеются четыре варианта для наращения процентов:

1. с конверсией: СКВ→Грн.→Грн.→СКВ;
2. с конверсией: Грн.→СКВ→СКВ→ГРН;
3. без конверсии: Грн.→Грн.;
4. без конверсии: СКВ→СКВ.

1. Валютные средства конвертируются в гривну, денежные средства в гривне размещаются в депозиты, наращение идет по ставке в гривнах, затем наращенная в гривнах сумма конвертируется обратно в исходную валюту.

2. Сумма в гривнах конвертируется в валюту и инвестируется в валютный депозит. Проценты начисляются по валютной ставке. Наращенная сумма в конце операции вновь конвертируется в гривны.

3. Сумма в гривнах инвестируется в депозит, на который начисляются проценты по ставке в гривнах (по формуле простых процентов).

4. Валютные средства размещаются в качестве валютного депозита по валютной ставке путем применения формул простых процентов.

В операции наращения с конверсией валют существует два источника дохода: изменение курса и начисление процента, причем, начисление процента является безусловным источником (ставка процента фиксирована, инфляция не рассматривается). Двойное конвертирование валюты может быть как источником дохода, так и приводить к убыткам.

Остановимся на двух вариантах, предусматривающих двойную конверсию.

Введем обозначения:

P_v — сумма депозита в СКВ;

P_g — сумма депозита в гривнах;

S_v — наращенная сумма в СКВ;

S_g — наращенная сумма в гривнах;

K_0 — курс обмена в начале операции (курс СКВ в грн.);

K_1 — курс обмена в конце операции;

n — срок депозита;

i — ставка наращения для депозитов в гривнах;

j — ставка наращения для конкретного вида СКВ.

1.7.1. Вариант СКВ→Грн.→Грн.→СКВ. (простые проценты)

Операция предполагает три этапа: обмен валюты на гривны, наращение процентов на эту сумму и обратное конвертирование в исходную валюту$\left[21\right]$. Наращенная сумма в валюте составит

$$S_v = P_v K_0 \left(1+ni\right)/K_1. \qquad (1.7.1)$$

Три этапа операции соответствуют полученной формуле. Множитель наращения с учетом двойного конвертирования равен

$$m = \frac{K_0}{K_1}\left(1+ni\right) = \frac{1+ni}{K_1/K_0} = \frac{1+ni}{K}, \qquad (1.7.2)$$

где $K = \dfrac{K_1}{K_0}$ − темп роста обменного курса за срок операции.

Из (1.7.2) следует, что множитель наращения m связан линейной зависимостью со ставкой i и обратной с обменным курсом в конце операции K_1 (или с темпом роста обменного курса K).

Исследуем зависимость общей доходности операции с двойной конверсией от темпа роста обменного курса K.

В качестве измерителя доходности операции в целом примем простую ставку процентов $i_{эфф}$ Пусть эта ставка характеризует рост суммы P_v до S_v. Для ее расчета воспользуемся формулой (1.7). Теперь ее можно записать как

$$i_{эфф} = \frac{S_v - P_v}{P_v n}.$$

Подставим в эту формулу значение S_v, полученное по (1.7.1). Тогда получим

$$i_{эфф} = \frac{\left(K_0/K_1\right)\left(1+ni\right)-1}{n} = \frac{1}{K}\frac{(1+ni)}{n} - \frac{1}{n}. \qquad (1.7.3)$$

Из (1.7.3) следует, что с увеличением K доходность $i_{эфф}$ падает. При K = 1 доходность операции равна ставке в гривнах $i_{эфф}$ = i, при K > 1 $i_{эфф}$ <

i, наконец, при $K < 1$ имеем $i_{эфф} > i$. Найдем теперь *критическое значение* K (обозначим его как K *), при котором $i_{эфф} = 0$. Из равенства (1.7.3) находим: $K^* = 1 + ni$, что означает

$$K_1^* = K_0(1+ni). \qquad (1.7.4)$$

Таким образом, если ожидаемые величины K или K_1 превышают свои критические значения, то операция убыточна.

Определим максимально допустимое значение курса обмена, при котором эффективность будет равна существующей ставке по депозитам в СКВ и применение двойного конвертирования не дает никакой дополнительной выгоды. Для этого приравняем множители наращения для двух альтернативных операций

$$1 + nj = (1+ni)K_0 / K_1.$$

Из последнего равенства следует, что

$$\max K_1 = K_0 \frac{1+ni}{1+nj}; \quad \max K = \frac{1+ni}{1+nj}.$$

Следовательно, депозит валюты через конверсию в гривны выгоднее валютного депозита, если обменный курс в конце операции будет меньше $\max K_1$.

1.7.2. Вариант Грн.→СКВ→СКВ→Грн (простые проценты)

Рассмотрим операцию с двойной конверсией, когда исходная и конечная суммы выражены в гривнах.

В этом случае трем этапам операции соответствуют три сомножите-ля для наращенной суммы

$$S_g = \frac{P_g}{K_0}(1+nj)K_1 = P_g(1+nj)\frac{K_1}{K_0}. \qquad (1.7.5)$$

Множитель наращения линейно зависит от процентной ставки j в валюте. Зависимость множителя от конечного курса или его темпа роста также линейная.

Проведем анализ эффективности данной операции и определим критические точки. Доходность операции определяется по формуле

$$i_{эфф} = \frac{S_g - P_g}{P_g n}.$$

Отсюда, подставив выражение S_g, получим

$$i_{эфф} = \frac{(K_1/K_0)(1+nj)-1}{n} = \frac{K(1+nj)-1}{n}. \qquad (1.7.6)$$

Зависимость показателя эффективности от K линейная. При K =1 $i_{эфф} = j$, при K > 1 $i_{эфф} > j$, при K < 1 $i_{эфф} < j$. Найдем критическое значение K *, при котором $i_{эфф} = 0$

$$K^* = 1/(1+nj); K_1^* = K_0/(1+nj).$$

Следовательно, если ожидаемые величины K или K_1 меньше своих критических значений, то операция убыточна.

Минимально допустимая величина K, обеспечивающая такую же доходность, что и прямой вклад в гривнах, определяется приравниванием множителей наращения для альтернативных операций (или из равенства $i_{эфф} = i$).

$$K_1(1+nj)/K_0 = 1+ni,$$

откуда $\min K = (1+ni)/(1+nj)$ или $\min K_1 = K_0(1+ni)/(1+nj)$.

Значит, депозит денежных сумм в гривнах через конверсию в валюту выгоднее депозита в гривнах, если обменный курс в конце операции больше $\min K_1$.

РАЗДЕЛ 2. НАРАЩЕНИЕ И ДИСКОНТИРОВАНИЕ ДЕНЕЖНЫХ СУММ ПО СЛОЖНЫМ ПРОЦЕНТНЫМ СТАВКАМ

2.1. ФОРМУЛА НАРАЩЕНИЯ СЛОЖНЫХ ПРОЦЕНТОВ

Второй метод наращения процентов заключается в том, что база начисления увеличивается ($S_1 = P + iP$), проценты начисляются на наращенную сумму ($P + iP$), что к концу второго периода начисления дает $S_2 = (P + iP) + i(P + iP) = (1+i)^2 P$, и т.д.

К концу n-го периода начисленная сумма станет равной

$$S_n = (1+i)^n P. \qquad (2.1.1)$$

В этом случае используются сложные проценты. Формула наращения сложных процентов имеет вид (2.1.1). Рассматривается финансовая операция, когда процентные деньги не платятся после каждого периода начисления, а присоединяются (реинвестируются) к основной сумме. Это означает, что база начисления меняется на каждом этапе: наращенная сумма S_n возрастает на долю i от предыдущей суммы S_{n-1}. Здесь $(1+i)^n -$ множитель наращения сложных процентов или мультиплицирующий множитель.

Наращение по сложным процентам представляет собой рост по закону геометрической прогрессии. Последовательность наращенных сумм $P, S_1, S_2, ..., S_n$ есть геометрическая прогрессия с начальным членом P и знаменателем прогрессии *(1+i)*.

Пример 1.

Годовая ставка сложных процентов равна 8%. Через сколько лет начальная сумма удвоится?

Решение:

Требуется решить неравенство: $(1+0{,}08)^n \geq 2$. Логарифмируем неравенство по основанию натуральных логарифмов и получаем $n \geq \ln 2 / \ln 1{,}08$.

Ответ: через 9 лет.

Формула наращения сложных процентов (2.1.1), выведенная для целых положительных n, может применяться и для нецелых t.

Сумма S_t, наращенная по ставке i сложных процентов, через t промежутков начисления станет

$$S_t = P(1+i)^t. \qquad (2.1.2)$$

Пример 2.

15 января в банк положили сумму 2000 *грн.* до востребования под ставку 12% годовых сложных процентов. Какую сумму снимет вкладчик 1 сентября?

Решение:

Применим формулу наращения сложных процентов (2.1.2). Будем считать, что в году 360 дней, в квартале — 90, в одном месяце — 30 и т.д. Тогда t =(30$\cdot$7+15)/360 и искомая сумма будет 2146,80 грн.

2.1.1..Формула удвоения наращенной суммы

При прогнозировании своих инвестиционных возможностей в будущем кредитор или должник могут задаться вопросом: по истечении скольких лет сумма ссуды возрастет в N раз при данной процентной ставке. Очевидно, что для ответа на этот вопрос следует приравнять множитель наращения величине N:

а) для простых процентов: $1+ni=N$, откуда $n=(N-1)/i.$

б) для сложных процентов: $(1+i)^n=N,$ откуда $n=\ln N/\ln(1+i)$.

Если $N=2$, то полученные формулы называются *формулами удвоения* и принимают следующий вид:

а) для простых процентов: $n=1/i.$

б) для сложных процентов: $n=\ln 2/\ln(1+i).$

Пример 3.

Вычислить, за сколько лет долг увеличится вдвое при ставке простых и сложных процентов равной 10%.

Решение.

а) для простых процентов: $n=1/0{,}1=10$ (лет).

б) для сложных процентов:

$n=\ln 2/\ln(1+0{,}1)=0{,}6932/0{,}0953=7{,}27$ (лет).

При использовании сложных процентов для приближенного оценивания полезно следующее *правило 72*.

Удвоение капитала по сложной процентной ставке α происходит примерно через $72/\alpha$ лет.

Например, согласно этому правилу при ставке 8%, удвоение капитала происходит за 9 лет.

2.2. СРАВНЕНИЕ СИЛЫ РОСТА ПРОСТЫХ И СЛОЖНЫХ ПРОЦЕНТОВ

Убедимся, что при одной и той же ставке i наращение сложных процентов будет быстрее, чем простых процентов, при длине периода наращения больше единичного и медленнее при длине периода наращения меньше единичного. Действительно,

$$(1+i)^t > 1+it, \text{ если } t>1,$$

$$(1+i)^t < 1+it, \text{ если } 0<t<1.$$

Графики функций $(1+i)^t$ и $(1+it)$ в зависимости от t представлены на рис. 2.1 [29, 30]

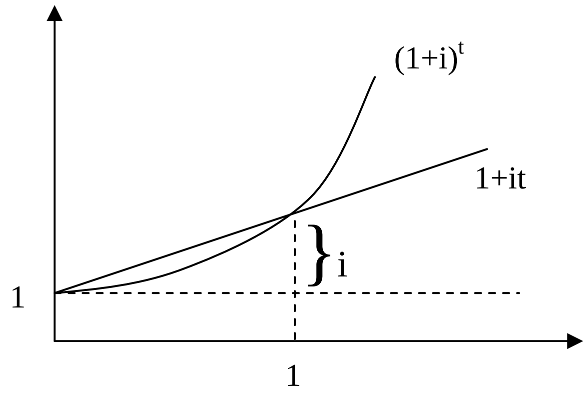


Рис.2.1

Отметим, что при сроке t<1 наращение по простым процентам дает больший результат, чем по сложным, а при сроке t>1 – наоборот. В этом нетрудно убедиться и на конкретных числовых примерах. Наибольшее превышение суммы, наращенной по простым процентам, над суммой, наращенной по сложным процентам, (при одинаковых процентных ставках) достигается в средней части периода начисления.

2.3. ФОРМУЛА НАРАЩЕНИЯ ПО СЛОЖНЫМ ПРОЦЕНТАМ ПРИ ПЕРЕМЕННОЙ ПРОЦЕНТНОЙ СТАВКЕ

Если период наращения n состоит из интервалов n_m, которым отвечают различные сложные процентные, ставки i_m, то наращенную сумму можно найти с помощью формулы

$$S = P(1+i_1)^{n_1}(1+i_2)^{n_2}...(1+i_m)^{n_m} = P\prod_{k=1}^{m}(1+i_k)^{n_k},$$

где $i_1, i_2, ...i_m$ – последовательные значения ставок процентов, действующих в периоды $n_1, n_2, ...n_m$ соответственно.

Пример 1.

При инвестировании 10 тыс. грн в депозит в договоре зафиксирована переменная ставка сложных процентов, определяемая как 20% годовых плюс маржа 10% в первые два года, 8% в третий год, 5% в четвертый год. Определить величину наращенной суммы за 4 года.

Решение.

$S = 10(1+0{,}3)^2(1+0{,}28)(1+0{,}25)=27{,}04$ (тыс. грн).

Пример 2.

Пусть сумма 1000 грн наращивается по ставке i =10% простых и сложных процентов. Тогда множители наращения $M_n^{(1)}$ и $M_n^{(3)}$ для простых и сложных процентов соответственно, а также наращенные суммы будут следующие:

n	0	0,5	1	10	100
$M_n^{(1)} = 1 + ni$ **S_n**	1 1000	1,05 1050	1,1 1100	2 2000	11 11000
$M_n^{(3)} = (1+i)^n$ **S_n**	1 1000	1,049 1049	1,1 1100	2,594 2594	13781 13781000

2.4. МНОЖИТЕЛИ НАРАЩЕНИЯ И ДИСКОНТИРУЮЩИЕ МНОЖИТЕЛИ

Множитель наращения для сложных процентов показывает во сколько раз возрастёт за n лет сумма, положенная в банк под i % годовых:

$$M_n^{(3)} = (1+i)^n.$$

Величина $M_n^{(3)}$ равна наращенной стоимости одной денежной единицы через n лет при ставке процента i. Например, $M_{10}^{(3)} = 2{,}594$; ($i = 0{,}1 = 10\%$, $n = 10$). Как правило, для облегчения расчетов составляются таблицы множителей наращения. Наличие электронных калькуляторов позволяет без затруднений выполнить необходимые расчеты. Однако, во многих ситуациях (например, в коммерческих банках) такие таблицы весьма удобны.

Дисконтирующие множители определяются по формуле

$$D_n^{(3)} = 1/M_n^{(3)} = (1+i)^{-n}.$$

Величину $D_n^{(3)}$ называют еще *приведенной или современной стоимостью* одной денежной единицы через n лет при сложной ставке процента i. Например, при i =10 имеем $D_{10}^{(3)} =$=0,386.

2.5. НАРАЩЕНИЕ ПРОЦЕНТОВ m РАЗ В ГОДУ. ЭФФЕКТИВНАЯ И НОМИНАЛЬНАЯ ПРОЦЕНТНЫЕ СТАВКИ

На практике начисление процентов может осуществляться чаще, чем 1 раз в год, например, раз в полугодие, квартал, месяц, неделю и даже сутки. Формула (2.1.1) остается справедливой и в этих случаях. Например, можно использовать такие ставки: «1% за каждую неделю» или «0,5% за квартал». Однако такой способ начисления процентов не является распространенным. На практике вместо реальной ставки i используют годовую ставку j, но указывают количество периодов начисления в течение года.

Пусть проценты начисляются m раз в год, а j − годовая ставка сложных процентов. Ставка j называется *номинальной*. На протяжении каждого из m периодов начисление процентов осуществляется по ставке j/m, тогда за год исходная сумма возрастет в $\left(1+\dfrac{j}{m}\right)^m$ раз, а за n лет в $\left[\left(1+\dfrac{j}{m}\right)^m\right]^n$ раз. Итак, для наращенной суммы получим

$$S = P\left[\left(1+\frac{j}{m}\right)^m\right]^n. \qquad (2.5.1).$$

Если $m = 1$, то будет получена формула (2.1.1).

Пример 1.

Вычислить наращенную сумму, если инвестировано 2000 грн. на 2 года. Используется сложная процентная ставка, равная 12% годовых с начислением процентов: 1) ежеквартально; 2) ежемесячно.

Решение.

1) $S = 2000\left(1 + 0{,}12 \cdot \frac{1}{4}\right)^{4\cdot 2} = 2533{,}54$ (грн), $m = 4$

2) $S = 2000 \cdot \left(1 + 0{,}12 \cdot \frac{1}{12}\right)^{12\cdot 2} = 2539{,}47$ (грн), $m = 12$.

Если $m = 1$, то наращенная сумма равна

$$S = 2000 \cdot (1 + 0{,}12)^2 = 2508{,}80 \text{ (грн)}.$$

Данный пример, является иллюстрацией того, что наращенная сумма увеличивается с возрастанием параметра m, то есть, чем чаще начисляют-ся проценты, тем быстрее идет процесс наращения.

Из формулы (2.5.1) получили формулы для вычисления срока наращения n и процентной ставки j:

$$n = \frac{\ln(S/P)}{m \cdot \ln\left(1 + \frac{j}{m}\right)}, \quad j = m\left(\sqrt[m\cdot n]{\frac{S}{P}} - 1\right) \qquad (2.5.2).$$

Пример 2.

По истечении какого времени сумма в 5000 грн увеличится до 5500 грн, если используется сложная процентная ставка j = 10%, а начисления выполняются 2 раза в год?

$$n = \frac{\ln(5500/5000)}{2 \cdot \ln\left(1 + \frac{0{,}1}{2}\right)} = 0{,}977 \text{ (лет)}.$$

Если срок ссуды измеряется дробным числом периодов начисления, то при m-разовом начислении процентов в году наращенную сумму можно вычислить по смешанной формуле:

$$S = P\left(1 + \frac{j}{m}\right)^k \left(1 + l\frac{j}{m}\right),$$

где $k-$ целое число периодов начисления ($k=\left[N/\tau\right]-$целая часть от деления всего срока ссуды N на период начисления τ), $l-$дробная часть периода начисления ($l=N/\tau-k$).

Формула (2.5.1) связывает пять параметров $S,P,m.n,j$. Тогда для определения каждого из них необходимо знать остальные четыре параметра. Однако часто рекламируются некорректные финансовые предложения, в которых приводятся лишь два и даже один параметр вместо четырех, необходимых для количественной оценки. Такая реклама рассчитана на непрозрачные условия кредитования. В США принят даже специальный закон, который требует от каждого кредитора вычисления так называемой *эффективной годовой процентной ставки*.

Эффективной ставкой — называется ставка, которая обеспечивает финансовый результат, эквивалентный тому, что соответствует номинальной ставке j с начислением процентов m раз в году. Очевидно, что эквивалентность финансовых результатов означает равенство $\left(1+i\right)^{n}=\left(1+\dfrac{j}{m}\right)^{mn}$, откуда эффективная ставка i равна

$$i=\left(1+\frac{j}{m}\right)^{m}-1. \qquad (2.5.3)$$

Отметим, что замена номинальной ставки j при m-разовом начислении процентов на эффективную ставку i не изменяет финансовых обязательств участвующих сторон, т.е. обе ставки эквивалентны в финансовом отношении.

Пример 3.

Вычислить эффективную ставку процента, если проценты начисляются ежеквартально, исходя из номинальной ставки 20% годовых.

$$i=\left(1+\frac{0,2}{4}\right)^{4}-1=0,2155\text{, т.е. } 21,55\%.$$

Пример 4.

Банк А предлагает инвестировать средства на срок 18 месяцев с использованием ставки 12% годовых и ежемесячным начислением процентов.

Банк Б предлагает инвестировать средства на срок 12 месяцев с использованием ставки 18% годовых, но с капитализацией процентов ежеквартально.

Какое предложение выгоднее инвестору?

Сравним эффективные годовые процентные ставки. Пользуясь формулой (2.5.3), находим

$$i_A = \left(1 + \frac{0{,}12}{12}\right)^{12} - 1 = 0{,}1268 \quad \text{или } 12{,}68\%,$$

$$i_B = \left(1 + \frac{0{,}18}{4}\right)^{4} - 1 = 0{,}1925 \text{ или } 19{,}25\%,$$

то есть, большее наращение обеспечивает предложение банка Б.

Из формулы (2.5.3) в случае необходимости можно решить обратную задачу — определить *j* по заданным значениям *i* и *m*. Находим

$$j = m\left(\sqrt[m]{1+i} - 1\right).$$

2.6. ДИСКОНТИРОВАНИЕ С ИСПОЛЬЗОВАНИЕМ СЛОЖНОЙ ПРОЦЕНТНОЙ СТАВКИ

В главе 1 значение настоящей (современной) величины P определенной денежной суммы S находилось с использованием простых процентов.

Рассмотрим схему определения величины P, используя сложные проценты. Итак, необходимо, исходя из величины денежной суммы в будущем S, найти величину денежной суммы P в произвольный предыдущий момент времени.

При изучении простых процентов для решения поставленной задачи рассматривались два метода: математическое дисконтирование и банковский (коммерческий) учет. Первый метод заключался в определении P по значению S при заданной ставке процента (математическое дисконтирование); второй — при заданной учетной ставке (банковский или коммерческий учет).

Применим математическое дисконтирование по сложной ставке процента. В соответствии с формулой (2.1.1) находим:

$$P=\frac{S}{(1+i)^n}=S(1+i)^{-n}=SD_n^{(3)}, \qquad (2.6.1)$$

где
$$D_n^{(3)}=(1+i)^{-n}=\frac{1}{M_n^{(3)}}. \qquad (2.6.2)$$

Величину $D_n^{(3)}$ называют *дисконтным множителем*. Значения множителя легко вычислить даже на калькуляторе.

Для случаев, когда проценты начисляются m раз в году, разрешая равенство (2.5.1) относительно P, получим:

$$P=\frac{S}{(1+j/m)^{mn}}=SD^{mn}, \qquad (2.6.3)$$

где
$$D^{mn}=(1+j/m)^{-mn}. \qquad (2.6.4)$$

Величину P, полученную дисконтированием S, называют *современной (настоящей) величиной* или *современной стоимостью* S. Современная стоимость находится на любой момент до выплаты суммы S. Разность S - P, в случае когда P определено дисконтированием, называют *дисконтом* и обозначают через D:

$$D=S-P=S(1-D_n^{(3)})$$

или при начислении процентов m раз в году $D=S-P=S(1-D^{mn})$.

Пример 1.

Найти современную величину долга в сумме 10000 грн, который надо уплатить через 3 года, если сложная процентная ставка — 20% годовых. Вычисления сделать для следующих моментов времени:

а) за $2,5$ года до конца срока,

б) за 1 год до конца срока.

Решение. Согласно формуле (2.6.1) получим

$$P=\frac{10000}{(1+0,2)^{2,5}}=6341,15 \text{ (грн)}, \qquad P=\frac{10000}{(1+0,2)^{1}}=8333,33 \text{ (грн)}.$$

Следовательно, сумма 6341,15 грн эквивалентна сумме 10000 грн за 2,5 года до конца срока, а сумма 8333,33 грн — за 1 год.

Пример 2.

Найти современную величину долга в сумме 20000 грн, который надо уплатить через 2 года. Рассмотреть три варианта уплаты процентов:

а) раз на год по годовой ставке 24%

б) раз на квартал по годовой ставке 12%

в) раз на месяц по годовой ставке 10%

Решение.

а) $P = \dfrac{20000}{\left(1+\dfrac{0,24}{1}\right)^{1\cdot 2}} = 13007,28$ (грн),

б) $P = \dfrac{20000}{\left(1+\dfrac{0,12}{4}\right)^{4\cdot 2}} = 15787,81$(грн),

в) $P = \dfrac{20000}{\left(1+\dfrac{0,10}{12}\right)^{12\cdot 2}} = 16388,20$ (грн).

2.7. ДИСКОНТИРОВАНИЕ И НАРАЩЕНИЕ ПО СЛОЖНОЙ УЧЕТНОЙ СТАВКЕ

Одной из важнейших характеристик, применяемых в финансовом анализе, является современная величина суммы денег.

Пусть S — дисконтируется сумма, P — ее современная стоимость, d — сложная годовая учетная (дисконтная) ставка. Тогда после первого дисконтирования получим

$P = S - Sd = S(1-d)$.

После второго — $P = S(1-d) - S(1-d)d = S(1-d)^2$.

После третьего — $P = S(1-d)^2 - S(1-d)^2 d = S(1-d)^3$.

Таким образом, дисконтируемые суммы образуют геометрическую прогрессию со знаменателем $q = 1-d$ и первым членом $b_1 = S$. По истечении n лет

$$P = S(1-d)^n. \qquad (2.7.1)$$

Таким образом, дисконтирование по сложной учетной ставке осуществляется по формуле (2.7.1).

В случае учета по сложной учетной ставке процесс дисконтирования происходит с замедлением, так как каждый раз учетная ставка применяется не к первоначальной сумме (как при простой учетной ставке), а к сумме, уже дисконтированной на предыдущем шаге во времени.

2.8. НОМИНАЛЬНАЯ И ЭФФЕКТИВНАЯ УЧЕТНЫЕ СТАВКИ

По аналогии с номинальной и эффективной процентной ставкой вводятся понятия "номинальная" и "эффективная учетная ставка". Обозначим номинальную учетную ставку как f. Если дисконтирование проводится m раз на протяжении года, то есть, в каждом периоде, равном $1/m$ части года, то дисконтирование осуществляется по сложной учетной ставке f/m. Процесс дисконтирования по данной сложной учетной ставке m раз в году описывается формулой

$$P = S\left(1-\frac{f}{m}\right)^{m\cdot n} = SD_{nm}^{(4)}, \tag{2.8.1}$$

где $f-$ номинальная годовая учетная ставка, $mn = N-$ общее число периодов дисконтирования, $D_{nm}^{(4)} = \left(1-\frac{f}{m}\right)^{m\cdot n}$ $-$дисконтный множитель.

2.8.1. Наращение по сложной учетной ставке

Ранее было рассмотрено наращение на основе сложной процентной ставки (ставки наращения). Однако, наращение достигается и с помощью сложной учетной ставки. Наращение является обратной задачей для учетных ставок. Формулы наращения по сложным учетным ставкам можно получить, разрешая соответствующие формулы для дисконтирования (2.7.1) и (2.8.1) Из формул (2.7.1) и (2.8.1) следует:

$$S = \frac{P}{(1-d)^n}; \quad S = \frac{P}{(1-f/m)^{mn}}. \tag{2.8.2}$$

Множитель наращения при использовании сложной ставки d, очевидно, равен $(1-d)^{-n}$.

2.8.2. Эффективная учетная ставка

Под эффективной учетной ставкой понимают сложную годовую учетную ставку, эквивалентную (по финансовым результатам) номинальной, применяемой при заданном числе m дисконтирований в году. В соответствии с определением эффективной учетной ставки можно найти ее связь с номинальной учетной ставкой из равенства дисконтных множителей

$$(1-d)^n = \left(1-\frac{f}{m}\right)^{mn}.$$

Эффективная учетная ставка характеризует результат дисконтирования за год и находится из равенства

$$1-d = \left(1-\frac{f}{m}\right)^{m}, \text{ откуда } \qquad d = 1-\left(1-\frac{f}{m}\right)^{m}.$$

Отметим, что эффективная учетная ставка всегда меньше номинальной.

Если дисконтная ставка меняется на протяжении периода n, то наращенную сумму можно вычислить с помощью формулы

$$S = \frac{P}{\prod_{i=1}^{m}(1-n_i d_i)},$$

где $n_1 + n_2 + ... + n_m = n$, а $d_1, d_2, ..., d_m$ – соответствующие ставки.

Пример 1.

Какую сумму следует проставить в векселе, если выданная сумма равна 50 тыс. грн., срок погашения 3 года. Вексель рассчитывается, исходя из сложной годовой учетной ставки 12%.

Решение. По формуле (2.9.2) находим

$$S = \frac{50}{(1-0{,}12)^3} = 73{,}371 \text{ (тыс.грн.)}$$

Пример 2.

Решить предыдущую задачу при условии, что наращение по сложной учетной ставке осуществляется ежеквартально.

Решение. Имеем $m = 4$, тогда

$$S = \frac{50}{\left(1-\frac{0{,}12}{4}\right)^{4\cdot 3}} = 72{,}063 \text{ (тыс.грн.)}$$

Пример 3.

Вычислить современную стоимость суммы в 20000 грн, что будет уплачена через два года. Сложная дисконтная ставка равна 10% и начисляется: а) ежегодно; б) ежеквартально.

Решение.

Согласно формулам (2.7.1) и (2.8.1), имеем

а) $P = 20000(1-0{,}1)^2 = 16200$ (грн).

б) $P = 20000\left(1-\dfrac{0{,}1}{4}\right)^{4\cdot 2} = 16333{,}04$ (грн).

Пример 4.

За два года до окончания операции дисконт равнялся третьей части долга. Вычислить соответствующую сложную дисконтную ставку.

Решение.

В соответствии с формулой (2.7.1) находим $P = S(1-d)^2.$

По условию задачи $D = S - P = S/3,$ отсюда $P = (2/3)S.$ Итак, $\dfrac{2}{3}S = S(1-d)^2,$ что дает $d = 1-\sqrt{\dfrac{2}{3}} = 0{,}184,$ или 18,4%.

2.9. НЕПРЕРЫВНОЕ НАРАЩЕНИЕ И ДИСКОНТИРОВАНИЕ

При использовании дискретных процентов наращенная сумма определяется по формуле (2.5.1)

$$S = P\left(1+\frac{j}{m}\right)^{mn},$$

где $j-$ номинальная процентная ставка, $m-$ число периодов начисления процентов в году. Промежутки времени между моментами начисления процентов уменьшаются с увеличением m.

Ранее были рассмотрены примеры зависимости скорости наращения от параметра m. Увеличение m, как мы видели, приводило также к возрастанию и наращенной суммы. Вполне естественно возникает вопрос о возможности дальнейшего наращения с увеличением m, то есть при условии, что промежутки времени между моментами начисления процентов будут неограниченно уменьшаться. Приведет ли это к неограниченному росту наращенной суммы? Исследуем формулу (2.5.1), откуда для множителя наращивания получим $\dfrac{S}{P} = \left(1+\dfrac{j}{m}\right)^{mn}$.

Исследование асимптотического поведения множителя наращения как функции m можно осуществить с помощью второго замечательного предела теории границ

$$\lim_{m\to\infty}\left(1+\frac{1}{m}\right)^m = e = 2,71828...$$

Тогда $\lim\limits_{m\to\infty}\left(1+\frac{j}{m}\right)^{mn} = \lim\limits_{m\to\infty}\left[\left(1+\frac{j}{m}\right)^{\frac{m}{j}}\right]^{jn} = e^{jn}$, где $e-$ основание натуральных логарифмов.

Используя полученный предел в выражении (2.5.1), находим, что наращенная сумма в случае непрерывного начисления процентов по ставке j будет равна

$$S = Pe^{jn}. \tag{2.9.1}$$

Непрерывным наращением по ставке j называется увеличение суммы в e^j раз за единичный промежуток начисления и в общем виде — увеличение суммы в e^{jn} раз за n промежутков начисления.

Пример. 10 млн. денежных единиц вложены на 4 года по ставке 80% годовых. Требуется найти наращенную за это время сумму с учетом простых и сложных процентов по годам. Найти наращенное значение вклада в конце 2-го года, если начисления производятся ежеквартально и ежемесячно. Сравнить результат, вычислив приращение исходной суммы.

Простые проценты. Пусть $S_к$ – наращенное значение вклада в конце k-го года. Для вычисления воспользуемся формулой $S_к=P(1+ki)$, где $i=\frac{80}{100}=0,8,$ тогда $S_к=10\cdot(1+0,8\cdot k)$; $k=\overline{1,4}.$

Результат представим в виде таблицы.

k	0	1	2	3	4	год
S_k	10	18	26	34	42	млн. ден. ед.

Сложные проценты. Воспользуемся формулой $S_k=P(1+i)^k$ для расчета наращенных значений вклада по годам. Тогда, по условиям задачи, $S_k=10(1+0,8)^k$; $k=\overline{1,4}$. Результаты представим в виде таблицы.

k	0	1	2	3	4	год
S_k	10	18	32,4	58,32	104,976	млн. ден. ед.

При одной и той же процентной ставке и первоначальной сумме вклада рост при вкладе под сложные проценты происходит значительно быстрее, чем при вкладе под простые проценты.

Найдем наращенное значение вклада по кварталам 2-го года, воспользовавшись формулой для равномерного начисления процентов: $S_n = P(1+\frac{i}{m})^{mn}$, где m=4 (начисление поквартально), n=2 (в конце 2-го года), i=0,8, P=10. Тогда $S_2 = 10(1+\frac{0,8}{4})^8 \cong 42,998.$

Найдем наращенное значение вклада в конце второго года при ежемесячном начислении процентов. В той же формуле теперь m=12, остальные значения – те же, тогда $S_2 = 10(1+\frac{0,8}{12})^{24} \cong 47,064.$

Для сравнения сведем результаты в таблицу, обозначив $\Delta = S_2 - P$ приращение первоначальной суммы при различных способах начисления**.**

начисление:	m	S_2 млн. ден. ед.	$\Delta = S_2 - P$
ежегодно	1	32,4	22,4
ежеквартально	4	42,998	32,998
ежемесячно	12	47,064	37,064

Для сравнения вычислим еще наращенное значение вклада по формуле непрерывного начисления процентов:

$S_n = P \cdot e^{i\,n}$; $S_2 = 10 \cdot e^{0,8 \cdot 2} = 10 \cdot e^{1,6} = 49,53$.

Для того, чтобы отличить ставку непрерывных процентов от ставок дискретных процентов, ее называют *силой роста* и обозначают символом δ. Тогда формула наращения примет вид

$$S = Pe^{\delta n}. \qquad (2.9.2)$$

Сила роста δ представляет собой номинальную ставку процента при условии, что $m \to \infty$.

Непрерывным дисконтированием называется операция, обратная непрерывному наращению, т.е. уменьшение суммы в e^j раз за единичный промежуток и уменьшение в e^{jn} раз за n промежутков. Выразив из (2.9.2) P через

S, получим формулу, по которой осуществляется дисконтиро-вание на основе непрерывных процентных ставок

$$P = Se^{-\delta n}. \tag{2.9.3}$$

2.10. СВЯЗЬ ДИСКРЕТНЫХ И НЕПРЕРЫВНЫХ ПРОЦЕНТНЫХ СТАВОК

Дискретные и непрерывные процентные ставки функционально зависимы, в силу чего можно осуществлять переход от расчета непрерывных процентов к дискретным процентам и наоборот. Формулу эквивалентного перехода от одних ставок к другим можно получить путем приравнивания соответствующих множителей наращения

$$(1+i)^n = e^{\delta n}. \tag{2.10.1}$$

Отсюда легко определить соответствующую эффективную ставку

$$i = e^{\delta} - 1, \tag{2.10.2}$$

то есть, наращение с использованием сложной процентной ставки i дает такой же результат, что и использование силы роста δ.

Из записанного равенства (2.10.2) также следует, что сила роста

$$\delta = \ln(1+i). \tag{2.10.3}$$

Пример.

Годовая ставка сложных процентов равна 20%. Найти эквивалент-ную силу роста.

Решение.

Используя формулу (2.10.3), получим $\delta = \ln(1+0,2) = 0{,}1823$, т.е. эквивалентная сила роста равна 18,23% и меньше годовой ставки сложных процентов, равной 20%.

Убедимся в том, что эффективная ставка i всегда будет больше номинальной (силы роста), то есть будет выполняться неравенство $i > \delta$. Действительно, воспользуемся разложением функции e^x в ряд Маклорена

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots + \frac{x^n}{n!} + \ldots \tag{2.10.4}$$

и полагая в формуле (2.10.4) $x = \delta$, находим

$$i = e^{\delta} - 1 = (1 + \delta + \frac{\delta^2}{2!} + \frac{\delta^3}{3!} + ...) - 1 = \delta + (\frac{\delta^2}{2!} + \frac{\delta^3}{3!} + ...).$$

Отсюда следует, что $i > \delta$, поскольку сила роста $\delta > 0$ и, следовательно, выражение в скобках (равное $i - \delta$) будет положительным.

2.11. НАХОЖДЕНИЕ СРОКА ССУДЫ И ПРОЦЕНТНЫХ СТАВОК

В некоторых практических задачах начальная сумма P и конечная сумма S могут быть известными по условиям контракта. Пусть требуется определить либо срок платежа, либо процентную ставку, которая в данном случае может служить характеристикой степени доходности финансовой операции для кредитора и мерой сравнения с рыночными показателями. Искомые величины могут быть определены из рассматриваемых ранее формул наращения или дисконтирования.

В практических финансовых соглашениях суммы денег связываются с некоторыми конкретными моментами или интервалами времени.

При разработке параметров соглашения фиксируются соответствующие даты, продолжительность операции (срок ссуды), остальные параметры сделки. Существуют различные способы начисления процентов, зависящие от условий контрактов. Соответственно применяются различные виды ставок. Используя ранее исследованные варианты расчета процентов, рассмотрим подробнее вопрос определения срока ссуды и процентных ставок.

а) При наращивании по сложной годовой ставке i используем формулу наращения $S = P(1+i)^n$ откуда следует, что

$$n = \frac{\ln(S/P)}{\ln(1+i)}, \qquad i = \left(\frac{S}{P}\right)^{1/n} - 1. \qquad (2.11.1)$$

б) При наращивании по номинальной ставке процентов m раз в году из формулы

$$S = P\left(1 + \frac{j}{m}\right)^{mn},$$

получаем:

$$n = \frac{\ln(S/P)}{m\ln(1 + j/m)}, \qquad j = m\left[\left(\frac{S}{P}\right)^{1/(mn)} - 1\right]. \qquad (2.11.2)$$

в) При дисконтировании по сложной годовой учетной ставке d. Используя формулу $P = S(1-d)^n$,находим:

$$n = \frac{\ln(P/S)}{\ln(1-d)}, \qquad d = 1 - \left(\frac{P}{S}\right)^{1/n}. \tag{2.11.3}$$

г) При дисконтировании по номинальной учетной ставке m раз в году. Из формулы $P = S\left(1 - \frac{f}{m}\right)^{m \cdot n}$ получим:

$$n = \ln(P/S)/(m\ln(1-f/m)), \qquad f = m\left(1 - \left(\frac{P}{S}\right)^{1/(mn)}\right). \tag{2.11.4}$$

д) При непрерывном наращении по постоянной силе роста из формулы $S = Pe^{\delta n}$ срок ссуды и сила роста соответственно равны:

$$n = \frac{\ln(S/P)}{\delta}, \qquad \delta = \frac{\ln(S/P)}{n}. \tag{2.11.5}$$

Пример.

За какой срок инвестированная сумма увеличится в два раза, если используется сложная дисконтная ставка $d = 20\%$ годовых, а начисления проводятся: а) ежеквартально; б) ежемесячно?

Решение.

В соответствии с формулой (2.11.4) получим:

а) $n = \dfrac{\ln 0{,}5}{4\ln(1-0{,}05)} = 3{,}38$ (лет),

б) $n = \dfrac{\ln 0{,}5}{12\ln(1-0{,}01666)} = 3{,}44$ (лет).

2.12. ВЛИЯНИЕ ИНФЛЯЦИИ НА ПРОЦЕНТНУЮ СТАВКУ ПО ПРОСТЫМ И СЛОЖНЫМ ПРОЦЕНТАМ

Ранее было отмечено, что деньги имеют временное преимущество, которое обусловлено, в частности, падением покупательной способности денег, то есть инфляцией. Например, если в начале года на 1 грн можно было купить 200 г сахара, то в начале следующего года только 180 г. Инфляцию невоз-можно игнорировать при анализе долгосрочных инвестиционных проектов.

Отметим, что анализ проблемы обесценения денег является сложной задачей, поскольку нуждается в учете многих экономических показателей. Однако во многих случаях можно ограничиться лишь вопросами корректировки процентной или дисконтной ставок. Считают, что инфляция (или уровень, темп инфляции) составляет долю h в год, если один и тот же набор товаров стоит в конце года в $(1+h)$ раз больше, чем в начале этого года. Это означает, что в $(1+h)$ раз уменьшилась покупательная способность одной денежной единицы. Как влияет инфляция на реальную ставку процента? Очевидно, что инфляция уменьшает реальную ставку процента. Это будет уже ставка процента с учетом инфляции. Действительно, одна денежная единица возрастает в течение года в $(1+i)$ раз в связи с наращением по процентной ставке i. Однако, ее покупательная способность уменьшается в $(1+h)$ раз, что обусловлено инфляцией. Таким образом, ее реальная ценность – покупательная способность – станет $(1+i)/(1+h)$, при этом реальная годовая ставка будет равной $(1+i)/(1+h)-1=(i-h)/(1+h)$. Видно, что при малой инфля-ции (когда h мало) реальная процентная ставка меньше номинальной приблизительно на величину инфляции. Для того чтобы номинальная ставка i обеспечивала наращение реальной ценности денежных сумм на долю j в год при годовой инфляции h, уровень инфляции должен удовлетворять уравнению: $(i-h)/(1+h)=j,$ отсюда $i=h+j(1+h)$. Остановимся подробнее на вопросах инфляционного обесценения денег и наращения по простым и сложным процентам.

Индекс цен (инфляции) − это отношение средневзвешенных цен определенного периода к средневзвешенным ценам базового периода (в США он рассчитывается на основе 265 товаров и услуг, которые входят в потребительскую корзину в 85 городах страны). Другими словами, индекс цен − это коэффициент роста цен за указанный промежуток времени.

Имеем
$$J=\frac{S}{S_0}, \tag{2.12.1}$$

где S – стоимость потребительской корзины в текущем периоде, S_0 – стоимость потребительской корзины в базовом периоде.

Уровень инфляции равен

$$h = \frac{\Delta S}{S_0} = \frac{S - S_0}{S_0}, \qquad (2.12.2)$$

то есть, $h = J - 1$, откуда:

$$J = 1 + h. \qquad (2.12.3)$$

Значит, индекс цен J показывает, во сколько раз увеличилась стоимость потребительской корзины, если уровень инфляции равен h.

Следствием инфляции является падение покупательной способности денег, которое за период n характеризуется индексом J_n.

Индекс покупательной способности равен обратной величине индекса цен J, т.е. $J_n = \frac{1}{J}$.

Пример 1.

Индекс цен в начале и в конце года составлял, соответственно, 1,25 и 1,39. Найти уровень инфляции.

Решение. В соответствии с формулой (2.12.2) находим

$$h = \frac{S_2 - S_1}{S_1} = \frac{S_0 \cdot 1,39 - S_0 \cdot 1,25}{S_0 \cdot 1,25} = 0,112,$$

или 11,2%, то есть цены возросли за год на 11,2 % или в 1,12 раз.

Покажем, что уровень инфляции не подчиняется аддитивному закону и что вычисление уровня инфляции осуществляется в соответствии с мультипликативным законом.

Для этого вычислим индексы инфляции в двух соседних периодах.

Пусть S_0 - стоимость потребительской корзины в начале первого периода, S_1 - стоимость корзины в конце первого периода, S_2 – соответствующая стоимость в конце второго периода.

Тогда индекс инфляции в первом периоде будет равен: $J_1 = \frac{S_1}{S_0}$, во втором периоде: $J_2 = \frac{S_2}{S_1}$.

Индекс инфляции за два периода составляет $J_{12} = \frac{S_2}{S_0}$, тогда имеем $J_1 \cdot J_2 = \frac{S_1}{S_0} \cdot \frac{S_2}{S_1} = \frac{S_2}{S_0}$, то есть $J_{12} = J_1 \cdot J_2$.

Используя равенство (2.12.3), получим: $1+h_{12}=(1+h_1)(1+h_2)$, откуда

$$h_{12}=h_1+h_2+h_1h_2, \tag{2.12.4}$$

то есть $h_{12}\neq h_1+h_2$.

Таким образом, уровень инфляции за несколько периодов не подчиняется аддитивному закону.

Обобщая формулу (2.12.4) для n периодов, можно показать, что будет справедлива формула

$$J_{12\ldots n}=(1+h_1)(1+h_2)\ldots(1+h_n)=\prod_{k=1}^{n}(1+h_k) \tag{2.12.5}$$

Если уровень инфляции постоянный $(h_1=h_2=\ldots=h_n)$, то

$$J=J_{12\ldots n}=(1+h)^n. \tag{2.12.6}$$

Пример 2.

Цены на протяжении 3 месяцев возрастали, соответственно за месяцами, на 5, 10, и 12 %. Найти уровень и индекс инфляции за три месяца.

Решение.

В соответствии с формулой (2.12.5) имеем

$$J=J_{123}=(1+0,05)(1+0,1)(1+0,12)=1,294,$$

отсюда $h=J-1=0,294$, или 29,4%, но не 5+10+12=27(%).

Пример 3.

Среднемесячный уровень инфляции равняется 10 %. Найти уровень инфляции за год.

Решение.

Согласно формуле (2.12.6) имеем индекс цен за год

$$J=J_{12\ldots 12}=(1+0,1)^{12}=3,1384.$$

Уровень инфляции за год составляет $h=J-1=2,1384$, или 213,84%, но не $0,1\cdot 12=1,2$, или 120%.

На практике реальная прибыль от инвестиции будет зависеть от инфляционных процессов и даже отсутствовать, поскольку инфляционные процессы и процессы наращения действуют в противоположных направлениях. Один из способов компенсации обесценения денег состоит в увеличении ставки процента на величину так называемой *инфляционной премии.* Скорректированная таким образом ставка называется брутто-ставкой. *Брутто-ставка*, которую мы будем обозначать символом r, находится из равенства скоррек-тированного на инфляцию множителя наращения по брутто-ставке множителю наращения по реальной ставке процента.

Формула Фишера для вычисления реальной прибыли использует индекс цен для корректировки ставки процента на величину инфляционной премии.

Пусть i — процентная ставка, h — уровень инфляции, S — наращен-ная сумма, C — наращенная сумма с учетом покупательной способности. Тогда для простых процентов

$$C = \frac{S}{J} = \frac{P(1+ni)}{(1+h)^n}. \tag{2.12.7}$$

Очевидно, что увеличение наращенной суммы (2.31) с учетом ее инфляционного обесценения имеет место только при условии $1+ni > J$.

Для сложных процентов

$$C = P\left(\frac{1+i}{1+h}\right)^n \tag{2.12.8}$$

Наращенная сумма (2.12 8) возрастает, если ставка i больше уровня инфляции h. Сумма не меняется, если $i = h$. Наращенная сумма уменьшается, если ставка i меньше уровня инфляции h.

Пусть $r-$ брутто-ставка, то есть, процентная ставка, учитывающая инфляцию. Тогда должно выполняться равенство

$$S = P(1+r).$$

Поскольку $$S = P(1+i)J = P(1+i)(1+h),$$

то $$P(1+r) = P(1+i)(1+h).$$

Следовательно, множитель наращения по брутто-ставке должен равняться множителю наращения по реальной ставке процента

$$(1+r) = (1+i)(1+h)$$

и брутто-ставка вычисляется по *формуле Фишера*:

$$r = i + (h + ih). \tag{2.12.9}$$

Слагаемое $(h+ih)-$ инфляционная премия, компенсирующая обесценение денег. Отметим, что ставка $(i+h)$ не компенсирует инфляцию, так как в сумме (2.12.9) есть еще слагаемое ih. Приближенная формула $r = i + h$ применяется только в случае, если величиной ih можно пренебречь.

2.13. ВЫЧИСЛЕНИЕ РЕАЛЬНОЙ СТАВКИ ПРОЦЕНТА

При наращении по простым процентам имеем равенство множителей наращения

$$1+nr=\left(1+ni\right)J, \tag{2.13.1}$$

где $J-$ индекс цен за период n. На практике требуется определять реальную доходность финансовой операции, т.е. доходность с учетом инфляции. Из соотношения (2.13.1) реальная ставка при начислении простых процентов равна

$$i=\frac{1}{n}\left(\frac{1+nr}{J}-1\right). \tag{2.13.2}$$

Реальная ставка при начислении сложных процентов из соотношения (2.12.9) находится по формуле:

$$i=\frac{r-h}{1+h}. \tag{2.13.3}$$

Пример 1.

Депозит в сумме 10000 грн размещен сроком на 2года. Ожидаемый уровень инфляции — 10 %. Сложная процентная ставка $i=15\%$. Вычислить наращенную сумму: а) без учета инфляции; б) с учетом инфляции.

Решение.

Согласно формулам (2.1.1) и (2.12.8) имеем

$S_2=P\left(1+i\right)^2=10000\left(1+0{,}15\right)^2=13225$ (грн).

$C=P\left(\dfrac{1+i}{1+h}\right)^n=10000\left(\dfrac{1+0{,}15}{1+0{,}1}\right)^2=10929{,}75$ (грн).

Пример 2.

Кредит в сумме 20000 грн выдано на 3 года. Ожидаемый уровень инфляции — 5 %. Определить простую процентную ставку кредита с учетом инфляции; величину наращенной суммы, если прибыльность операции должна составлять 10% годовых.

Решение.

В соответствии с формулой (2.12.6) индекс цен за период $n=3$ равен

$J=\left(1+h\right)^3=\left(1+0{,}05\right)^3=1{,}158.$

Из формулы (2.13.1) находим $r=\dfrac{\left(1+ni\right)\left(1+h\right)^n-1}{n},$

что дает $$r = \frac{(1+3\cdot 0{,}1)1{,}158-1}{3} = 0{,}1685,$$ то есть 16,85%.

При наращении по простым процентам с учетом инфляции для наращенной суммы получим $C = P(1+rn)$.

Таким образом, наращенная сумма равна

$$C = 20000(1+0{,}1685\cdot 3) = 30110.$$

2.14. КОНВЕРСИЯ ВАЛЮТЫ: ВАРИАНТ СКВ→ГРН.→ГРН.→СКВ. (СЛОЖНЫЕ ПРОЦЕНТЫ)

Рассмотрим совмещение конвертирования валюты в гривны, наращение сложных процентов в гривнах и затем конвертирование полученной суммы в гривнах в исходную валюту [21]. Наращенная сумма, получаемая в конце операции, состоящей из трех указанных этапов, будет равна

$$S_v = P_v K_0 (1+i)^n \frac{1}{K_1},$$

где $P_v -$ сумма депозита в валюте,

$S_v -$ наращенная сумма в валюте,

$K_0 -$курс обмена в начале операции (курс валюты в грн.),

$K_1 -$ курс обмена в конце операции,

$i -$ставка сложных процентов в гривнах,

$j -$ ставка наращения для конкретной валюты.

Множитель наращения с учетом двойного конвертирования равен

$$m = (1+i)^n \frac{K_0}{K_1} = \frac{(1+i)^n}{k}, \qquad (2.14.1)$$

где $k = \dfrac{K_1}{K_0} -$ темп роста валютного курса за срок операции.

Множитель наращения m связан обратной зависимостью с обменным курсом в конце операции K_1 (или с темпом роста обменного курса k).

Исследуем зависимость общей доходности операции от соотношения конечного и начального курса k. Доходность операции в целом определим в виде годовой ставки сложных процентов $i_э$.

Используем формулу наращения (2.1.1) по сложным процентам $S = P(1+i)^n$, из которой следует, что $i_э = \sqrt[n]{\frac{S_v}{P_v}} - 1.$

Подставив в эту формулу S_v из (2.14.1), получим

$$i_э = \sqrt[n]{\frac{P_v (1+i)^n K_0 / K_1}{P_v}} - 1 = \frac{1+i}{\sqrt[n]{k}} - 1. \qquad (2.14.2)$$

Очевидно, что с увеличением темпа роста k эффективность $i_э$ падает.

Анализируя зависимость (2.14.2) получаем, что при $k = 1$ $i_э = i,$ при $k > 1$ $i_э < i$, а при $k < 1$ $i_э > i.$

Значение k, при котором эффективность операции равна нулю, т.е. критическое значение, определяется из равенства $i_э = 0$. Отсюда $k_{kp} = (1+i)^n$, что означает равенство среднегодового темпа роста курса валюты годовому темпу наращения по ставке в гривнах: $\sqrt[n]{k} = 1 + i.$

Таким образом, если ожидаемая величина k больше своего критического значения, то операция с двойным конвертированием убыточна, так как $i_э < 0$. Максимально допустимое значение $k_{\max}$, при котором доходность операции будет равна доходности при прямом инвестировании валютных средств по ставке j, находится из равенства соответствующих множителей наращения

$$\frac{(1+i)^n}{k_{\max}} = (1+j)^n.$$

Отсюда $k_{\max} = \left(\frac{1+i}{1+j}\right)^n$ или $K_{1\max} = K_0 \left(\frac{1+i}{1+j}\right)^n.$

Конвертация валюты в гривны выгоднее валютного депозита, если обменный курс в конце операции будет меньше $K_{1\max}$.

РАЗДЕЛ 3. КРЕДИТНЫЕ РАСЧЕТЫ

3.1. КРЕДИТНЫЕ ОПЕРАЦИИ

Одной из основных финансовых операций является *кредитная сделка*. Кредит, ссуда – древнейшие финансовые операции. По-латыни «creditum» означает «ссуда». «Кредит», «ссуда» − это предоставление денег или товаров в долг на условиях возвратности и, как правило, с уплатой процентов через точно определенный срок за их использование. Условия выдачи и погашения кредитов весьма разнообразны: открытие сберегательного счета в банке, выдача банком кредита, учет векселя и др. Юридическим обеспечением кредитной сделки служит финансовый контракт между *кредитором* — лицом, предоставляющим в долг финансовые средства и *дебитором* (заемщиком, должником) — лицом, получающим финансовые средства для временного их использования. С позиции кредитора сущность кредитной сделки состоит в получении определенной выгоды, дохода от сделанной им инвестиции. Для дебитора процент представляет собой стоимость кредита и рассматривается им как убытки или издержки.

3.2. ПРОСТАЯ КРЕДИТНАЯ СДЕЛКА

Простая кредитная сделка связывает две суммы: величину выданного кредита P и его полную (вместе с процентами I) стоимость S; при этом очевидно, что

$$S = P+I. \tag{3.2.1}$$

Равенство (3.2.1) отражает финансовую сущность простейшей кредит-ной сделки и называется *основной формулой теории кредитных операций.*

Простая кредитная сделка представляет собой единовременную выдачу займа, погашаемого одним платежом в конце срока сделки.

Пусть кредит P выдан на n лет под i сложных годовых процентов. К концу n-го года наращенная сумма будет равна $P(1+i)^n$. Если требуется отдать заем одним платежом, то это и есть размер данного платежа.

Пример.

Заем величиной 10000 грн. был выдан на 5 лет под 10% годовых. Вычислить размер долга, если предполагается вернуть этот заем одним платежом.

Решение.

По формуле $S = P(1+i)^n$ находим

$S = 10000(1+0{,}1)^5 = 16105{,}10$ (грн) − искомый платеж.

Сумма выданного кредита называется основным долгом, а наращиваемый добавок – процентными деньгами. Пусть заем P выдан на n лет под i годовых процентов. За 1-й год процентные деньги составят iP. После выплаты процентных денег останется снова только основной долг в размере P. Пусть наращенные за истекший год процентные деньги iP выплачиваются в конце каждого года. В конце n-го, последнего года выплаты составят величину $P + iP -$ основной долг и процентные деньги за последний год.

3.3. ПОТРЕБИТЕЛЬСКИЙ КРЕДИТ

С развитием рыночных отношений стало возможным предоставление потребительских кредитов. Это относится, прежде всего, к приобретению ценных товаров: бытовой техники, автомобилей, мебели и др. Существуют различные схемы погашения потребительского кредита. При выдаче потребительского кредита сразу на всю сумму кредита начисляются простые проценты, которые прибавляются к величине самого кредита и все погашающие выплаты будут равны полученной величине.

Рассмотрим две простейших схемы такого кредитования.

3.3.1. Погашение кредита равными выплатами

Потребитель получает товар ценой P в кредит сроком на n лет под i простых годовых процентов. Следовательно, сумма выплат по формуле (1.4.1) будет составлять

$$S = P(1+ni).$$

Если в договоре о кредите предусмотрено m выплат в год равными частями, то каждая часть будет равна

$$h = P(1+ni)/nm.$$

Пример:

Холодильник стоимостью 3500 грн. продано в кредит на 2 года. Процентная ставка 20% годовых. Погашение кредита проводится ежемесячно. Какую сумму покупатель уплатит магазину за весь период, если:

а) кредит выдан на всю сумму;

б) 25% стоимости товара покупатель уплатил в начале операции?

Решение.

Согласно формуле (1.7) имеем

а) $S = 3500(1 + 2 \cdot 0,2) = 4900$ (грн), то есть цена телевизора фактически увеличится на 4900−3500=1400(грн). Ежемесячная плата будет составлять

$$h = \frac{4900}{2 \cdot 12} = 204,17 \text{ (грн)}.$$

Если часть платы продавец получил в момент передачи товара, то проценты начисляются на остаток задолженности. Тогда для второй схемы имеем:

б) $S = 3500 \cdot 0,75(1 + 2 \cdot 0,2) = 2625 \cdot 1,4 = 3675$ (грн),

то есть цена телевизора возрастет на (3675−2625)=1050(грн), а еже-месячная плата будет равна

$$h = \frac{3675}{2 \cdot 12} = 153,13 \text{ (грн)}.$$

3.3.2. Правило 78

На практике часто применяется схема, когда кредит выплачивается равными долями, а проценты в размере $niP -$ выплатами, уменьшающимися в арифметической прогрессии, то есть начисляющимися на остаток долга. Последняя выплата равна разности этой прогрессии. Таким образом, разовая выплата составляется из двух слагаемых — соответствующая часть основного долга (уплата) и процентная выплата, которая постоянно уменьшается. Пусть в год планируется m выплат (при ежемесячных выплатах $m = 12$), тогда последняя выплата равна $d -$ разность прогрессии, а первая − mnd. Сумма всех выплат будет равна $d + 2d + ... + mnd = d(1 + mn)mn / 2$ и должна быть равна процентным деньгам niP. Используя равенство $d(1 + mn)mn / 2 = niP$, можно определить d и ежемесячные выплаты.

На практике можно вычислить сумму номеров всех выплат $M = 1 + 2 + ... + mn = (1 + mn)mn / 2$ и разделить процентные деньги на M

частей. Тогда первый платеж будет равен mn таких частей, второй − на одну часть меньше и т.д., последний − одной части. Название правила объясняется тем, что сумма номеров месяцев в году равна 1+2+…+12=78.

На практике удобно также поступить следующим образом.

Пусть $P-$ сумма кредита, которая выплачивается в течение m месяцев, а $i-$ годовая процентная ставка. Тогда основная сумма P погашается равными частями

$$h = P/m. \tag{3.3.1}$$

Процентная плата в первый месяц составляет $h_1 = Pi/12$. Тогда месячная выплата равна $h + h_1 = P(1/m + i/12)$.

Для вычисления суммы, которая будет выплачена за весь период (m месяцев), необходимо определить сумму всех процентных платежей:

за второй месяц $\qquad h_2 = (P - P/m)\cdot\dfrac{i}{12} = P(1-1/m)\cdot\dfrac{i}{12},$

за третий месяц $\qquad h_3 = (P - 2P/m)\cdot\dfrac{i}{12}.$

Тогда за k-й месяц процентная выплата будет равна

$$h_k = \left(P - (k-1)P/m\right)\cdot\frac{i}{12}, k \le m. \tag{3.3.2}$$

Сумма процентных денег за m месяцев:

$$I = \sum_{k=1}^{m} h_k = \left(1 + \left(1-\frac{1}{m}\right) + \left(1-\frac{2}{m}\right) + \ldots\right.$$

$$\left.\ldots + \left(1-\frac{k-1}{m}\right) + \ldots + \left(1-\frac{m-1}{m}\right)\right)\frac{Pi}{12} =$$

$$= \left(m - \frac{1}{m}(1+2+3+\ldots+m-1)\ \right)\frac{Pi}{12}.$$

Суммируя арифметическую прогрессию в скобках, находим сумму процентных денег за весь период

$$I = \frac{Pi}{24}(m+1). \tag{3.3.3}$$

С учетом погашения основной суммы долга, получим величину, которая будет выплачена в течение m месяцев

$$S = P + I = P\left(1 + \frac{i}{24}(m+1)\right). \tag{3.3.4}$$

Пример 3.3.

Для сопоставления результатов рассмотрим еще раз пример (3.3.1) при кредитовании по правилу 78.

Решение. Согласно формуле (3.3.3) получим

$$I = \frac{3500 \cdot 0,2}{24} \cdot (24+1) = 729,17 \text{ (грн)},$$

то есть фактическая цена холодильника увеличится не на 1050 грн, а на 729,17 грн и будет равняться, согласно (3.3.4), S = 3500+729,17 = 4229,17 (грн). Средняя величина ежемесячной платы тоже будет меньшей: $h = (P+J)/m = 4229,17/24 = 176,22$ (грн).

Приведенные примеры показывают, что выбор схемы кредитования может существенно повлиять на финансовые результаты. В рассмотренных случаях начисляется одинаковая процентная ставка и одинаковый период начисления, но результаты существенно отличаются. Причина заключается в том, что в последней схеме база начисления процентов все время уменьшается.

3.4. ЭКВИВАЛЕНТНОСТЬ В ФИНАНСОВЫХ РАСЧЕТАХ

В финансовых расчетах возникает необходимость в сравнении результатов операции, то есть в сравнении сумм, отвечающим различным моментам времени. Результаты конкретных финансовых операций зависят от вида и величины ставки наращения или дисконтирования, величины исходной суммы, периода начисления.

В простейшем случае, если ставки применяются к равным исходным суммам $P_1 = P_2$ равенство $S_1 = S_2$ или $P_1 k_1 = P_2 k_2$ переходит в равенство для множителей наращения $k_1 = k_2$.

Суммы S_1 и S_2 называются *эквивалентными*, если их величины, вычисленные в определенный момент времени с помощью одинаковой ставки, совпадают. Вычисление этих сумм выполняется с помощью формул наращения или дисконтирования в зависимости от выбора момента времени.

Пусть надо сравнить сумму S_1 (срок уплаты через n_1 лет) и сумму S_2 (срок уплаты через n_2 лет). Пусть также выполняются неравенства $S_1<S_2$ и

$n_1<n_2$. Для сравнения этих сумм можно вычислить их современные значения с помощью, например, сложной процентной ставки, то есть сравнить

$$P_1 = \frac{S_1}{(1+i)^{n_1}} \text{ и } P_2 = \frac{S_2}{(1+i)^{n_2}}.$$

При малых i ,очевидно, имеем $P_1 \approx S_1, P_2 \approx S_2$, следовательно, $P_1 = P_2$.

При достаточно больших i знак неравенства изменится на противоположный, т.е. $P_1 > P_2$. Отсюда следует существование такого значения величины процентной ставки i_{kp}, для которого выполняется равенство

$$\frac{S_1}{(1+i)^{n_1}} = \frac{S_2}{(1+i)^{n_2}}, \qquad (3.4.1)$$

Откуда

$$i_{kp} = \left(\frac{S_2}{S_1}\right)^{\frac{1}{n_2-n_1}} - 1. \qquad (3.4.2)$$

Значит, $P_1 < P_2$, если $i < i_{kp}$; $P_1 = P_2$.,если $i = i_{kp}$; $P_1 > P_2$, если $i > i_{kp}$.

При использовании простой процентной ставки вместо равенства (3.4.1) будем иметь

$$\frac{S_1}{1+n_1 i_{kp}} = \frac{S_2}{1+n_2 i_{kp}}, \qquad (3.4.3)$$

Откуда

$$i_{kp} = \frac{S_2 - S_1}{S_1 n_2 - S_2 n_1}. \qquad (3.4.4)$$

Пример:

Заемщик, который должен возвратить 15000 грн через 2 года, предлагает кредитору 18000 грн через 3 года. Выгодно ли для кредитора данное предложение, если простая процентная ставка составляет 12%?

Для сравнения указанных сумм приведем их к одному моменту времени с помощью дисконтирования. Вычислим современную стоимость сумм:

$$P_1 = \frac{S_1}{1+n_1 i} = \frac{15000}{1+2\cdot 0{,}12} = 12096{,}77,$$

$$P_2 = \frac{S_2}{1+n_2 i} = \frac{18000}{1+3\cdot 0{,}12} = 13235{,}29.$$

Поскольку $P_1 < P_2$, предложение заемщика является выгодным для кредитора.

Для сравнения воспользуемся формулой (3.4.4), что дает:

$$i_{kp} = \frac{18000 - 15000}{15000 \cdot 3 - 18000 \cdot 2} = 0,33 \text{ или } 33\%.$$

Итак, $P_1 < P_2$, т.к. i =12% < 33% < i_{kp}.

Если в последнем примере воспользоваться сложными процентами, то по формуле (3.4.2) получим:

$$i_{kp} = \left(\frac{S_2}{S_1}\right)^{\frac{1}{n_2 - n_1}} - 1 = \left(\frac{1800}{1500}\right)^{\frac{1}{3-2}} - 1 = 0,2 \text{ или } 20\%.$$

Итак, $i < i_{kp}$, следовательно, $P_1 < P_2$. В этом случае также приемлемо предложение заемщика.

Рассмотрим теперь сравнение сумм S_1 и S_2 при использовании дисконтной ставки. В соответствии с формулой (1.5.1) имеем:

$$P_1 = \frac{S_1}{1 - d_1 n_1},\ P_2 = \frac{S_2}{1 - d_2 n_2}.$$

Определим d_{kp} из уравнения $\frac{S_1}{1 - d_{kp} n_1} = \frac{S_2}{1 - d_{kp} n_2}$, откуда: $d_{kp} = \frac{S_1 - S_2}{S_1 n_2 - S_2 n_1}$.

При использовании сложной дисконтной ставки критическое значение находим из уравнения $\frac{S_1}{(1+i)^{n_1}} = \frac{S_2}{(1+i)^{n_2}}$, откуда: $d_{kp} = 1 - \left(\frac{S_2}{S_1}\right)^{\frac{1}{n_2 - n_1}}$.

3.5. ОБЪЕДИНЕНИЕ ЗАЙМОВ

При заключении финансовых сделок между заемщиком и кредитором существует эквивалентность финансовых обязательств. Очевидно, что изменение срока оплаты или величины некоторых из платежей может изменить окончательный финансовый результат в пользу одной из сторон. Используя идею эквивалентности, можно несколько займов объединить в один. Для этого сначала определяют современные величины остатков займов, затем эти величины складывают. В соответствии с принципом эквивалентности финансовых результатов компенсация изменений параметров финансовой операции должна осуществляться таким образом, чтобы не нарушались интересы участников. В каждом конкретном случае требуется составить уравнение эквивалентности, в котором все денежные суммы приведены к одному моменту времени.

Сначала рассмотрим задачу консолидации платежей. Пусть $S_1, S_2, ..., S_N$ – суммы, подлежащие объединению, а $n_1, n_2, ..., n_N$ – соответствующие моменты времени. В соответствии с принципом эквивалентности необходимо, чтобы величина денежной суммы S_0 в момент n_0 равнялась сумме всех платежей $S_1, S_2, ..., S_N$, приведенных к тому же моменту n_0.

Если $n_0 > n_1, n_2, ..., n_N$, то уравнение эквивалентности будет иметь вид:

$$S_0 = \sum_{k=1}^{N} S_k (1 + it_k), \tag{3.5.1}$$

где $t_k = n_0 - n_k$.

Если $n_0 < n_1, n_2, ..., n_M$, то уравнение эквивалентности будет иметь вид:

$$S_0 = \sum_{m=1}^{M} S_m (1 + it_m)^{-1}, , \tag{3.5.2}$$

где $t_m = n_m - n_0$.

В общем случае будем иметь уравнение

$$S_0 = \sum_{k=1}^{N} S_k (1 + it_k) + \sum_{m=1}^{M} S_m (1 + it_m)^{-1}, \tag{3.5.3}$$

где первая сумма соответствует платежам, произведенным до момента n_0, а вторая сумма − платежам, произведенным после момента n_0.

При решении конкретных примеров удобно пользоваться так называемыми временными диаграммами. Временная диаграмма − это горизонтальная ось, на которой цифрой 0 обозначается настоящий момент времени, а цифрами *1, 2,…* − моменты времени *1, 2, …*периода. Периодом может быть год, квартал, месяц, день и т.д.

Денежные суммы размещаются над номерами периодов, а процентные ставки − над соответствующими отрезками временной оси.

Рис 3.1 иллюстрирует финансовую операцию, состоящую в размещении на депозите 1000 грн в начале операции, изъятии 200 грн через 1 период и внесении 300 грн еще через один период, X – сумма, которая накапливается через 3 периода.

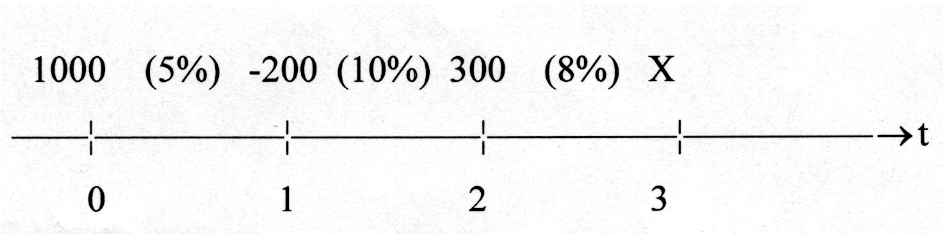


Рис 3.1

Задачи консолидации платежей с использованием сложной процентной ставки отличаются лишь соответствующим множителем наращения, а именно, вместо уравнений (3.5.1) − (3.5.3) будем иметь следующие уравнения эквивалентности:

$$S_0 = \sum_{k=1}^{N} S_k (1+i)^{t_k}, \text{ если } n_0 > n_1, n_2, ..., n_N.$$

$$S_0 = \sum_{k=1}^{M} S_m (1+i)^{-t_m}, \text{ если } n_0 > n_1, n_2, ..., n_m.$$

В общем случае

$$S_0 = \sum_{k=1}^{N} S_k (1+i)^{t_k} + \sum_{k=1}^{M} S_m (1+i)^{-t_m}. \qquad (3.5.4)$$

Пример:

Три платежа 1000, 1500 и 2000 грн со сроками платежей 1, 2 и 4 года консолидируются в один через 3 года. Найти соответствующую сумму, если используется сложная процентная ставка $i = 5\%$.

Очевидно, что первые две суммы необходимо наращивать, а последнюю − дисконтировать. Рис 3.2 иллюстрирует данную финансовую операцию.

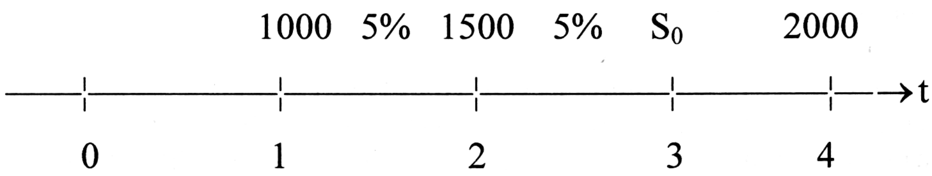


Рис 3.2

В соответствии с формулой (3.5.4) получим

$$S_0 = 1000(1+0,05)^{3-1} + 1500(1+0,05)^{3-2} + +2000(1+0,05)^{-(4-3)} = 4582,26 \text{ (грн)}.$$

РАЗДЕЛ 4. ФУНКЦИИ МНОГИХ ПЕРЕМЕННЫХ В ЭКОНОМИКЕ

4.1. ИССЛЕДОВАНИЕ ФУНКЦИИ МНОГИХ ПЕРЕМЕННЫХ НА ЭКСТРЕМУМ

Пусть функция n независимых переменных $u = f(x_1, x_2, \ldots, x_n) = f(M)$ определена в некоторой δ-окрестности точки $M_0\left(x_1^0, x_2^0, \ldots, x_n^0\right)$, то есть при $|MM_0|<\delta$, где δ — сколь угодно малое положительное число.

Функция $u = f(M)$ имеет в точке M_0 максимум, если существует δ-окрестность точки M_0, для всех точек которой выполняется неравенство $f(M) \le f(M_0)$; функция u имеет минимум, если $f(M) \ge f(M_0)$.

Максимум или минимум называются ее экстремумами.

4.1.1. Необходимые условия существования экстремума

Если функция $u = f(x_1, x_2, \ldots, x_n)$ в точке M_0 имеем экстремум, то в этой точке либо все частные производные 1-го порядка обращаются в 0, либо хотя бы одна из них не существует, а остальные равны 0.

Точки, в которых выполняются необходимые условия существования экстремума функции, называются *критическими.*

Точки, в которых частные производные существуют и равны нулю $\frac{\partial u}{\partial x_i} = 0\left(i = \overline{1,n}\right)$, называются *стационарными* точками.

Полученные критические точки исследуются на экстремум с помощью достаточных условий экстремума.

4.1.2. Достаточные условия существования экстремума

Пусть функция $u = f(x_1, x_2, \ldots, x_n)$ имеет частные производные второго порядка в окрестности стационарной точки M_0.

Это означает, что в данной точке M_0 существует дифференциал второго порядка.

$$d^2u\Big|_{M_0} = \sum_{i=1}^{n}\sum_{k=1}^{n}\frac{\partial^2 u}{\partial x_i \partial x_k}\Bigg|_{M_0} dx_i dx_k.$$ Обозначим: $a_{ik} = \frac{\partial^2 u}{\partial x_i \partial x_k}\Big|_{M_0}$; $a_{ik} = a_{ki}\ \left(i,k = \overline{1,n}\right)$.

Таким образом, $d^2u\Big|_{M_0} = \sum_{i=1}^{n}\sum_{k=1}^{n} a_{ik}dx_i dx_k$, где a_{ik} — коэффициенты дифференциала второго порядка в точке M_0. Составим матрицу из элементов a_{ik} дифференциала $d^2u\Big|_{M_0}$:

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}.$$

Запишем главные матрицы

$$A_1 = a_{11},\ A_2 = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix},\ A_3 = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix},\ \dots,\ A_n = \begin{vmatrix} a_{11} & \dots & a_{1n} \\ \dots & \dots & \dots \\ a_{n1} & \dots & a_{nn} \end{vmatrix}.$$

По формуле Тейлора в окрестности точки M_0 имеем

$$f(M) = f(M_0) + df(M_0) + \frac{1}{2!}d^2 f(M)$$

Поскольку в стационарной точке M_0 $df(M_0) = 0$, знак разности $f(M) - f(M_0)$ определяется знаком второго дифференциала $d^2 f(M)$.

На основании критерия Сильвестра сформулируем *достаточные условия существования экстремума функции* $u = f(x_1, x_2, \dots, x_n)$.

1) Если все главные миноры матрицы A положительные, то есть $A_1 > 0, A_2 > 0, \dots, A_n > 0$, то функция $u = f(M)$ в точке M_0 имеет минимум.

2) Если знаки главных миноров чередуются, начиная с $A_1 < 0$, то есть $A_1 < 0, A_2 > 0, A_3 < 0, \dots$, то функция $u = f(M)$ в точке M_0 имеет максимум.

3) Если $d^2 u(M_0) = 0$, то требуются дополнительные исследования.

Если условия 1) или 2) не выполняются, то экстремума нет.

4.1.3. Исследование функции многих переменных на условный экстремум. Пусть функция $u = f(\text{M}) = f(x_1, x_2, \dots, x_n)$ рассматривается не во всей области определения**,** а только на некотором множестве E**,** координаты всех точек которого удовлетворяют уравнениям:

$$\begin{cases} \varphi_1(x_1, x_2, \dots, x_n) = 0 \\ \varphi_2(x_1, x_2, \dots, x_n) = 0 \\ \dots\dots\dots\dots\dots\dots\dots \\ \varphi_m(x_1, x_2, \dots, x_n) = 0 \end{cases},\ где\ \ m < n. \qquad (4.1.1)$$

Эти уравнения называют уравнениями связи. Предположим, что существует точка $M_0 \in E$ и такая окрестность $C_\varepsilon(M_0) \in E$, что для всех $M \in C_\varepsilon(M_0) \cap E$ будет выполняться $f(M) > f(M_0)$ ($f(M) < f(M_0)$). Тогда утверждают, что функция $u=f(M)$ имеет в точке M_0 условный минимум (условный максимум) при заданных условиях связи (4.1.1).

Для нахождения точек, подозреваемых на условный экстремум, строим вспомогательную функцию Лагранжа

$$F(x_1, x_2, \dots, x_n; \lambda_1, \lambda_2, \dots, \lambda_m)=f(x_1, x_2, \dots, x_n)+\sum_{k=1}^{m}\lambda_k\varphi_k\left(x_1, x_2 \dots x_n,\right)\lambda_k\varphi_k(x_1, x_2, \dots, x_n), \quad (4.1.2)$$

где λ_k $(k=\overline{1,m})$ – постоянные множители Лагранжа, и приравниваем нулю частные производные по всем ее $(n+m)$ переменным.

При этом знак второго дифференциала $d^2F(x_1^o, x_2^o, \dots, x_n^o)$ в стационарной точке M_0 $(x_1^o, x_2^o, \dots, x_n^o)$ определяет характер экстремума при условии, что переменные $dx_1, dx_2, \dots, dx_n$ связаны соотношениями

$$\left\{\sum_{k=1}^{n}\frac{\partial\varphi_j}{\partial x_k}dx_k=0,\quad j=\overline{1,m}\right.$$

***4.1.4. Метод исключения части переменных и сведения задачи к задаче на безусловный экстремум*:**

$$\begin{cases} x_1 = F_1\left(x_{k+1},\dots,x_n\right) \\ \dots\dots\dots\dots\dots\dots\dots\dots \\ x_k = F_k\left(x_{k+1},\dots,x_n\right) \end{cases}$$

Подставляя $x_1, \dots, x_k$ в выражение u, получим:

$$u = f[F_1(x_{k+1}, \dots, x_n), \dots, F_k(x_{k+1}, \dots, x_n), x_{k+1}, \dots, x_n] = \varphi(x_{k+1}, \dots, x_n) = \varphi(N).$$

Значение функции $u=\varphi(N)$ в любой точке N совпадает со значением $u=f(M)$ в соответствующей точке $M(x_1, x_2, \dots, x_n)$, удовлетворяющей уравнениям связи. Поэтому задача об условном экстремуме функции $u=f(M)$ сводится к задаче о безусловном экстремуме функции $u=\varphi(N)$.

Однако не всегда возможно разрешить уравнения связи относительно каких-либо "k" переменных, например, $x_1, x_2, \dots, x_k$ и получить в явном виде

$$\begin{cases} x_1 = F_1\left(x_{k+1},\dots,x_n\right) \\ \dots\dots\dots\dots\dots\dots\dots\dots\ . \\ x_k = F_k\left(x_{k+1},\dots,x_n\right) \end{cases}$$

В этом случае используется метод неопределенных множителей Лагранжа

4.1.5. Метод неопределенных множителей Лагранжа

Для нахождения точек, подозреваемых на условный экстремум, строим вспомогательную функцию Лагранжа

$$F(x_1, x_2, \dots, x_n; \lambda_1, \lambda_2, \dots, \lambda_m) = f(x_1, x_2, \dots, x_n) + \sum_{k=1}^{m} \lambda_k \varphi_k \left(x_1, x_2 \dots x_n,\right), \qquad (4.1.3)$$

где λ_k $(k = \overline{1, m})$ — постоянные множители Лагранжа, и приравниваем нулю частные производные по всем ее $(n + m)$ переменным.

При этом знак второго дифференциала $d^2F(x_1^o, x_2^o, \dots, x_n^o)$ в стационарной точке M_0 $(x_1^o, x_2^o, \dots, x_n^o)$ определяет характер экстремума при условии, что переменные $dx_1, dx_2, \dots, dx_n$ связаны соотношениями

$$\left\{ \sum_{k=1}^{n} \frac{\partial \varphi_j}{\partial x_k} dx_k = 0, \quad j = \overline{1,m} \right.$$

Пример 1. Исследовать на экстремум функцию

$z = \left(3x^2 - 4y\right)^2 + \frac{1}{3}x + \frac{2}{3}y$ при условии связи $\frac{3}{4}x^2 - y + \frac{1}{6} = 0.$

Решение.

Составим функцию Лагранжа

$$L = L(x, y, \lambda) = \left(3x^2 - 4y\right)^2 + \frac{1}{3}x + \frac{2}{3}y + \lambda\left(\frac{3}{4}x^2 - y + \frac{1}{6}\right)$$

Необходимые условия экстремума: $\frac{\partial L}{\partial x} = 0, \frac{\partial L}{\partial y} = 0, \quad \frac{\partial L}{\partial \lambda} = 0.$

$$\begin{cases} \frac{\partial L}{\partial x} = 12x\left(3x^2 - 4y\right) + \frac{1}{3} + \frac{3}{2}\lambda x \\ \frac{\partial L}{\partial y} = -8\left(3x^2 - 4y\right) + \frac{2}{3} - \lambda \\ \frac{\partial L}{\partial \lambda} = \frac{3}{4}x^2 - y + \frac{1}{6} \end{cases}$$

$$\begin{cases} 12x\left(3x^2 - 4y\right) + \frac{1}{3} + \frac{3}{2}\lambda x = 0 \\ -8\left(3x^2 - 4y\right) + \frac{2}{3} - \lambda = 0 \\ \frac{3}{4}x^2 - y + \frac{1}{6} = 0 \end{cases} \Rightarrow x_0 = -\frac{1}{3}, y_0 = \frac{1}{4}, \lambda = 6.$$

Точка $M_0\left(\frac{-1}{3};\frac{1}{4}\right)$ — точка возможного экстремума функции z при данном условии связи.

К функции Лагранжа применим достаточные условия экстремума. Находим частные производные от функции Лагранжа

$$\frac{\partial^2 L}{dx^2}=12\left(9x^2-4y\right)+\frac{3}{2}\lambda,$$
$$\frac{\partial^2 L}{\partial x\partial y}=-48x,$$
$$\frac{\partial^2 L}{\partial y^2}=32.$$

Вычислим полученные частные производные в точке M_0 при $\lambda = 6$.

$$a_{11}=\left.\frac{\partial^2 L}{\partial x^2}\right|_{M_0}=9,\;\; a_{12}=\left.\frac{\partial^2 L}{\partial x\partial y}\right|_{M_0}=16,\;\; a_{22}=\left.\frac{\partial^2 L}{\partial y^2}\right|_{M_0}=32.$$

Вычислим второй дифференциал d^2L в точке возможного экстремума:

$$\left.d^2L\right|_{M_0}=a_1dx^2+2a_{12}dxdy+a_{22}dy^2=9dx^2+32dxdy+32dy^2.$$

Знакоопределенность квадратичной формы $\left.d^2L\right|_{M_0}$ требуется установить при условии связи $\varphi(x,y)=\frac{3}{4}x^2-y+\frac{1}{6}=0,$ поэтому следует найти дифференциал $d\varphi$ и вычислить его в точке $M_о$. Имеем

$$\left.d\varphi\right|_{M_0}=\left.\frac{\partial\varphi}{\partial x}\right|_{M_0}dx+\left.\frac{\partial\varphi}{\partial y}\right|_{M_0}dy=\left.\frac{2}{3}x\right|_{M_0}dx-dy=-\frac{1}{2}dx-dy=0\Rightarrow$$
$$\Rightarrow dy=-\frac{1}{2}dx.$$

Подставим dy в $\left.d^2L\right|_{M_0}$, тогда:

$$\left.d^2L\right|_{M_0}=9\cdot dx^2+32\cdot dx^2\left(-\frac{1}{2}\right)+32\cdot dx^2\cdot\left(-\frac{1}{2}\right)^2=dx^2.$$

Поскольку $\left.d^2L\right|_{M_0}=dx^2>0,$ функция Лагранжа имеет минимум.

$L_{\min}$ = *1/2* $\Rightarrow$ $z = f(x, y)$ в точке $M_0\left(-\frac{1}{3};\frac{1}{4}\right)$ имеет $z_{\min}=\frac{1}{2},$ так как при наличии связи L и z совпадают.

4.2. ПРОИЗВОДНАЯ ПО НАПРАВЛЕНИЮ

Рассмотрим функцию $U = f(x, y, z)$ и найдем величину, характеризующую скорость ее изменения в некоторой точке $M_0(x_0, y_0, z_0)\in D$ в направлении единичного вектора $\vec{l} = (\cos\alpha; \cos\beta, \cos\gamma)$, где $\cos\alpha$, $\cos\beta$, $\cos\gamma$— направляющие косинусы вектора $\vec{l}$. Обозначим через $M(x, y, z)\in D$ переменную точку на луче l (Рис.4.1).

Определение. *Производной функции f(x, y, z) по направлению* $\vec{l}$ *в точке* M_0 *называется предел*

$$\frac{\partial f}{\partial l} = \lim_{M\to M_0}\frac{f(M)-f(M_0)}{|MM_0|}.$$

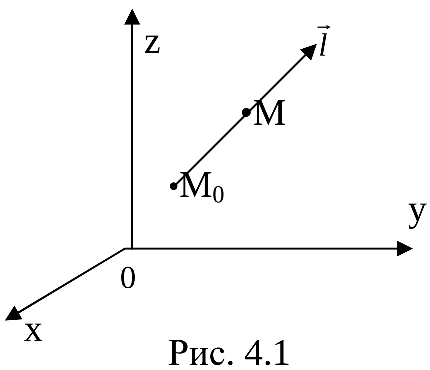


Рис. 4.1

Если направление $\vec{l}$ совпадает с положительным направлением одной из осей OX, OY или OZ, то $\frac{\partial f}{\partial l}$ — частная производная $\frac{\partial f}{\partial x}$, $\frac{\partial f}{\partial y}$ или $\frac{\partial f}{\partial z}$.

Теорема. *Если функция f(x, y, z) имеет в точке* M_0 *непрерывные частные производные (т.е. дифференцируема в точке* M_0*), то в точке* M_0 *существует производная по любому направлению, причем, эта производная определяется формулой*

$$\frac{\partial f}{\partial l} = \frac{\partial f}{\partial x}\cos\alpha + \frac{\partial f}{\partial y}\cos\beta + \frac{\partial f}{\partial z}\cos\gamma. \qquad (4.2.1)$$

Производная $\frac{\partial U}{\partial l}$ характеризует *скорость изменения величины U(M) в направлении* $\vec{l}$.

Пример 2.

Найти производную функции $U = xy^2z^3$ в точке $A(3; 2; 1)$ по направлению вектора $\vec{a} = 2\vec{i} + 2\vec{j} + \vec{k}$.

Решение.

Запишем формулу (4.2.1) для производной функции U по направлению $\vec{a}$ в точке A:

$$\frac{\partial U}{\partial a}=\left(\frac{\partial U}{\partial x}\cos\alpha+\frac{\partial U}{\partial y}\cos\beta+\frac{\partial U}{\partial z}\cos\gamma\right)\Bigg|_A$$

Частные производные

$$\left.\frac{\partial U}{\partial x}\right|_A=\left.y^2z^3\right|_A=4,\ \left.\frac{\partial U}{\partial y}\right|_A=\left.2xyz^3\right|_A=12,\ \left.\frac{\partial U}{\partial z}\right|_A=\left.3xy^2z^2\right|_A=36.$$

Направляющие косинусы вектора $\vec{a}$ равны: $\cos\alpha=\frac{2}{\sqrt{2^2+2^2+1^2}}=\frac{2}{3}$; $\cos\beta=\frac{2}{3}$, $\cos\gamma=\frac{1}{3}$, где $|\vec{a}|=3$; отсюда $\frac{\partial U}{\partial a}=4\cdot\frac{2}{3}+12\cdot\frac{2}{3}+36\cdot\frac{1}{3}=22\frac{2}{3}$.

4.3. ГРАДИЕНТ ФУНКЦИИ

Градиент функции — мера возрастания величины на единицу длины (латинск. gradiens).

Пусть функция $f(x, y, z)$ определена и дифференцируема в области D.

Определение 1. Вектор, проекции которого на координатные оси равны соответствующим частным производным функции $f(x, y, z)$, называется градиентом функции $f(x, y, z)$

$$\mathrm{grad} f=\frac{\partial f}{\partial x}\vec{i}+\frac{\partial f}{\partial y}\vec{j}+\frac{\partial f}{\partial z}\vec{k}.$$

Пользуясь понятием производной по направлению, получим некоторые свойства градиента.

СВОЙСТВА ГРАДИЕНТА

1. Перепишем формулу (4.2.1) для производной по направлению $\vec{l}$ в виде скалярного произведения

$$\frac{\partial f}{\partial l}=\left(\mathrm{grad}\, f,\vec{l}\right)=\left|\mathrm{grad}\, f\right|\left|\vec{l}\right|\cos\varphi=\left|\mathrm{grad}\, f\right|\cos\varphi,$$

где $\left|\vec{l}\right|=1$.

Итак, $\dfrac{\partial f}{\partial l} = |\mathrm{grad}\, f| \cos\varphi.$

Из этой формулы следует, что производная функции по направлению $\vec{l}$ имеет наибольшее значение, если $cos\varphi = 1$, т.е., если угол φ между векторами $grad\ f$ и $\vec{l}$ в точке M_0 равен 0. В этом случае направление $\vec{l}$ совпадает с направлением $grad\ f$.

Следовательно, если вектор $\vec{l}$ направлен вдоль градиента функции в данной точке, т.е. $\vec{l} \uparrow\uparrow \mathrm{grad}\, f$, то $\dfrac{\partial f}{\partial l} = |\mathrm{grad}\, f|$ и принимает наибольшее значение в данной точке.

Полученный результат позволяет теперь взамен приведенного выше формального определения градиента, связанного с выбором системы координат, дать другое определение, не зависящее от выбора системы координат.

Определение 2. *Градиентом функции U(M) называется вектор, направленный в сторону наибольшего возрастания функции и по модулю равный производной функции U по этому направлению*. Причем, это наибольшее значение равно $|\mathrm{grad}\, f| = \sqrt{\left(\dfrac{\partial f}{\partial x}\right)^2 + \left(\dfrac{\partial f}{\partial y}\right)^2 + \left(\dfrac{\partial f}{\partial z}\right)^2}$.

2. Направление градиента совпадает с направлением нормали к поверхности уровня $U(x, y, z) = C$, проходящей через данную точку.

3. Производная функции по любому направлению, касательному к поверхности уровня, равна 0.

4.4. НЕКОТОРЫЕ ПРИМЕНЕНИЯ ФУНКЦИЙ НЕСКОЛЬКИХ ПЕРЕМЕННЫХ В ЭКОНОМИКЕ

4.4.1. Графический метод нахождения максимума и минимума линейной функции в области. Использование понятия градиента

Найдем оптимальное (максимальное или минимальное) значение функции

$$f = c_1x_1 + c_2x_2 \qquad (4.4.1)$$

при ограничениях
$$\begin{cases} a_{11}x_1 + a_{12}x_2 \le b_1, \\ a_{21}x_1 + a_{22}x_2 \le b_2, \\ \dots\dots\dots\dots\dots\dots \\ a_{k1}x_1 + a_{k2}x_2 \le b_k \\ x_1 \ge 0, x_2 \ge 0 \end{cases} \qquad (4.4.2)$$

Каждое из неравенств системы ограничений задачи геометрически определяет полуплоскость соответственно с граничными прямыми

$$a_{i1}x_1 + a_{i2}x_2 = b_i \left(i = \overline{1,k}\right), x_1 = 0, x_2 = 0.$$

Если система неравенств (4.4.2) совместна, область D ее решений есть множество точек, принадлежащих всем полуплоскостям. Это множество называется многоугольником решений. Стороны этого многоугольника лежат на прямых, уравнения которых получается из исходной системы ограничений заменой знаков неравенств на знаки равенств. Таким образом, задача состоит в нахождении такой точки многоугольника решений, в которой функция f, называемая функцией цели (или целевой функцией), принимает оптимальное значение. Эта точка существует тогда, когда многоугольник решений не является пустым множеством.

Для определенности будем рассматривать задачу отыскания максимального значения внутри многоугольника D решений или на его границе.

Если бы f принимала максимальное значение внутри области D, то ее частные производные должны равняться нулю или не существовать.

В силу линейности функции $f = c_1x_1 + c_2x_2$ $\dfrac{\partial f}{\partial x_i} = c_i \ne 0, \left(i = 1, 2\right)$, откуда следует, что функция f может принимать свое максимальное значение на границе области D, причем, из приведенных соображений это могут быть вершины многоугольника решений (можно показать, что если f принимает свое максимальное значение в двух соседних вершинах многоугольника решений, то это значение она принимает и на отрезке, соединяющем данные вершины).

Для определения данной вершины строим линию уровня $c_1x_1 + c_2x_2 = h$, где h – некоторая постоянная, проходящая через многоугольник решений и будем перемещать ее в направлении вектора $\vec{c} = (c_1; c_2)$ до тех пор, пока она не пройдет через её последнюю общую точку с многоугольником решений. Координаты указанной точки и определяют оптимальное решение (план) данной задачи.

Напомним, что градиентом функции $f(x_1, x_2)$ в точке $\vec{x} = (x_1, x_2)$ называют вектор, координаты которого равны значениям соответствую-щих частных производных этой функции в данной точке, т.е.

$$\overrightarrow{\mathrm{grad}}\, f\,(\overline{x}) = \left(\frac{\partial f}{\partial x_1}; \frac{\partial f}{\partial x_2}\right) = (c_1; c_2) = \vec{c}.$$

Таким образом, в каждой точке плоскости $\overrightarrow{grad}\ f = \overrightarrow{const}$ и совпадает с вектором, составленным из коэффициентов данной функции при переменных. Очевидно, что $\overrightarrow{grad}\ f = \overrightarrow{grad}\,(c_1x_1 + c_2x_2) = c_1\vec{i} + c_2\vec{j} = \vec{c}$ перпендикулярен каждой линии ее уровня.

Предельное положение прямой, полученное с помощью её параллельного перемещения в направлении градиента (в направлении противоположном градиенту) целевой функции, отвечает ее максимальному (минимальному) значению.

Приведенные рассуждения являются основой графического метода решения задачи линейного программирования.

Если множество допустимых решений задачи не ограничено, то целевая функция может оказаться неограниченной.

Нахождение максимального оптимального решения задачи (4.4.1), (4.4.2) на основе её геометрической интерпретации включает следующие этапы:

1. Строят прямые, уравнения которых получаются в результате замены в ограничениях знаков неравенств на знаки равенств.
2. Находят полуплоскости, определяемые каждым из ограничений задачи.
3. Находят многоугольник решений.

4. Строят вектор $\vec{C} = (c_1, c_2) = \overrightarrow{\text{grad}}\, f$.
5. Строят прямую $c_1x_1 + c_2x_2 = h$, проходящую через многоугольник решений.
6. Передвигают прямую $c_1x_1 + c_2x_2 = h$ в направлении вектора $\vec{C}$, в результате чего либо находят точку (точки), в которой целевая функция принимает максимальное значение, либо устанавливают неограниченность сверху функции на множестве планов.
7. Определяют координаты точки максимума функции и вычисляют значение целевой функции в этой точке.

Пример 3.

Для производства двух видов изделий *A* и *B* предприятие использует три вида сырья. Нормы расхода сырья каждого вида на изготовление единицы продукции данного вида приведены в таблице 4.1. В ней же указаны прибыль от реализации одного изделия каждого вида и общее количество сырья данного вида, которое может быть использовано предприятием.

Таблица 4.1.

Вид сырья	Нормы расхода сырья (кг) на одно изделие		Общее количество сырья (кг)
	A	**B**	
1 2 3	12 4 3	4 4 12	300 120 252
Прибыль от реализации одного изделия (грн.)	30	40	

Учитывая, что изделия *A* и *B* могут производиться в любых соотношениях, требуется составить такой план выпуска, при котором прибыль предприятия от реализации всех изделий является максимальной.

Математическая модель задачи

Предположим, что предприятие изготовит x_1 изделий вида *A* и x_2 изделий вида *B*. Поскольку производство продукции ограничено имеющимся в распоряжении предприятия сырьем каждого вида и количество изготовляемых изделий не может быть отрицательным, должны выполняться неравенства

$$\begin{cases} 12x_1 + 4x_2 \le 300, \\ 4x_1 + 4x_2 \le 120, \\ 3x_1 + 12x_2 \le 252, \end{cases}$$
$$x_1, x_2 \ge 0$$

Общая прибыль от реализации x_1 изделий вида A и x_2 изделий вида B составит $f = 30x_1+40x_2$.

Таким образом, мы приходим к следующей математической задаче: среди всех неотрицательных решений данной системы линейных неравенств требуется найти такое, при котором функция f принимает максимальное значение.

4.4.2. Нахождение экстремума в задачах потребительского выбора

Постановка задачи потребителя: имеется n видов товаров и услуг, количества которых $x_1, x_2, \ldots, x_n$ по ценам $P_1, P_2, \ldots, P_n$ соответственно.

Суммарная стоимость этих товаров составляет $\sum_{i=1}^{n} P_i x_1$. Уровень потребления $U = f(x_1, x_2, \ldots, x_n)$ — функция полезности.

Необходимо найти такой набор товаров и услуг $x_1, x_2, \ldots, x_n$ при данной величине дохода I, чтобы обеспечить максимальный уровень потребления, т.е., чтобы $U = f(x_1, x_2, \ldots, x_n) \to max$ при условии $\sum_{i=1}^{n} P_i x_i \le I, x_i \ge 0 \left(i = \overline{1,n}\right)$. Функцией спроса называется решение этой задачи, зависящее от цен $P_1, P_2, \ldots, P_n$ и величины дохода I.

Сумма $\sum_{i=1}^{n} P_i x_i$. — сумма денег, затрачиваемая на приобретение товаров; $P_i x_i$ — на приобретение i-го товара.

Пример 5.

Решить задачу потребительского выбора, найдя функции спроса, при ценах благ $P_1 = 10$, $P_2=10$ и доходе $I = 60$ с функцией предпочтения:

$U = x_1^{\frac{1}{2}} x_2^{\frac{2}{3}} \to \max,$ где x_1, x_2 — количество благ первого и второго видов соответственно.

Следовательно, $x_1P_1 + x_2P_2 \le 60$ или $10x_1 + 10x_2 \le 60 \Rightarrow x_1 + x_2 \le 6$, $x_1 \ge 0$, $x_2 \ge 0$.

Решение:

Строим область допустимых значений x_1 и x_2. Для этого строим прямые: $x_1 = 0$, $x_2 = 0$, $x_1 + x_2 = 6$.

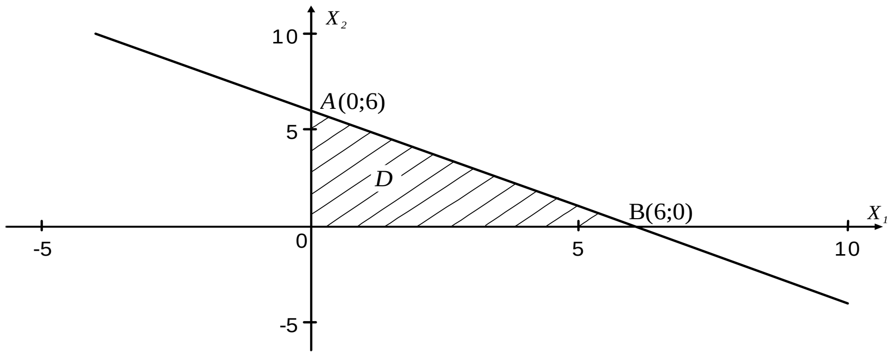


Рис. 4.2

Итак, требуется определить максимальное значение функции предпочтения U при условии, что x_1 и x_2 принадлежат области D.

Функция U принимает своё максимальное значение либо внутри области D, либо на её границе. Для нахождения максимального значения внутри области D находим критические точки.

$$\begin{cases} \dfrac{\partial U}{\partial x_1} = 0, \\ \dfrac{\partial U}{\partial x_2} = 0; \end{cases} \quad \begin{cases} \dfrac{\partial U}{\partial x_1} = \dfrac{1}{2} x_1^{-\frac{1}{2}} x_2^{\frac{2}{3}}, \\ \dfrac{\partial U}{\partial x_2} = \dfrac{2}{3} x_1^{\frac{1}{2}} x_2^{-\frac{1}{3}}; \end{cases} \Rightarrow \begin{cases} x_1^{-\frac{1}{2}} x_2^{\frac{2}{3}} = 0, \\ x_1^{\frac{1}{2}} x_2^{-\frac{1}{3}} = 0; \end{cases} \Rightarrow \begin{cases} x_1 \ne 0,\ x_2 = 0, \\ x_2 \ne 0,\ x_1 = 0. \end{cases}$$

Критические точки принадлежат границе области D. Следовательно, максимальное значение функции предпочтения U может находиться на границе области D.

На стороне OA: $U=0$; на стороне OB: $U=0$.

На стороне AB: $x_2=6-x_1$;

$$U/_{AB} = x_1^{\frac{1}{2}} (6 - x_1)^{\frac{2}{3}} = U(x_1).$$

Исследуем эту функцию на экстремум:

$$U'(x_1)=\frac{1}{2}x_1^{-\frac{1}{2}}(6-x_1)^{\frac{2}{3}}-x_1^{\frac{1}{2}}\cdot\frac{2}{3}(6-x_1)^{-\frac{1}{3}}=\frac{1}{x_1^{\frac{1}{2}}(6-x_1)^{\frac{1}{3}}}\left(\frac{6-x_1}{2}-\frac{2}{3}x_1\right)=\frac{18-7x_1}{6x_1^{\frac{1}{2}}(6-x_1)^{\frac{1}{3}}}.$$

$$U'(x_1)=0\Rightarrow x_1=\frac{18}{7};\ x_1\neq 0,\ x_1\neq 6.\quad x_2=6-\frac{18}{7}=\frac{24}{7};\quad M_1\left(\frac{18}{7};\frac{24}{7}\right)$$

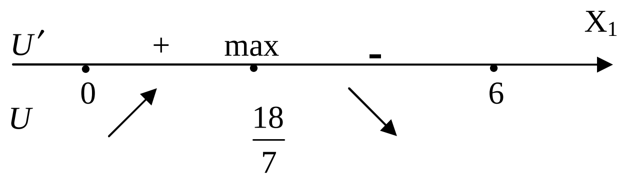


Рис.4.3

В точке M_1: $U\left(\frac{18}{7};\frac{24}{7}\right)=\left(\frac{18}{7}\right)^{\frac{1}{2}}\left(\frac{24}{7}\right)^{\frac{2}{3}}$;

В угловых точках: $U(0;0)=0,\ U(B)=U(6;0)=0;\ \ U(A)=U(0;\ 6)=0.$

Ответ: $U_{\max}\left(\frac{18}{7};\frac{24}{7}\right)=\left(\frac{18}{7}\right)^{\frac{1}{2}}\left(\frac{24}{7}\right)^{\frac{2}{3}};\quad x_1=\frac{18}{7};\ x_2=\frac{24}{7}.$

Пример 6.

Решить задачу потребительского выбора при тех же ценах: $P_1=10$, $P_2=10$, доходе $I=60$ со следующей функцией предпочтения

$U=3(x_1-2)^2+2(x_2-3)^2\rightarrow max,\ \ x_1\geq 0,\ x_2\geq 0.$

Решение.

Построим область D допустимых значений x_1 и x_2. Для этого построим прямые $x_1=0$, $x_2=0$; по условию $10x_1+10x_2\leq 60\Rightarrow x_1+x_2\leq 6$ (Рис.4.4).

Строим прямую $x_1+x_2=6$.

Итак, требуется найти максимальное значение функции предпочтения U при условии, что x_1 и x_2 принадлежат области D.

Функция U своё наименьшее значение равное 0, принимает в точке $O_1(2;\ 3)$.

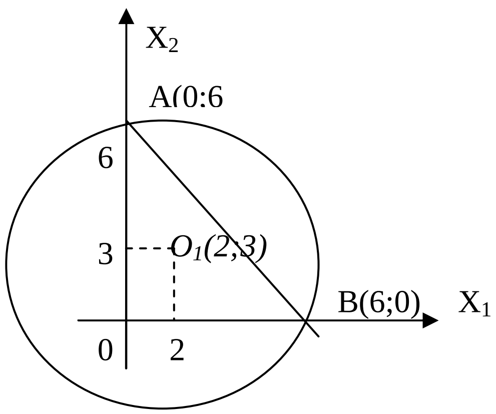


Рис.4.4.

С увеличением расстояния от точки O_1 функция возрастает и наибольшее значение принимает в наиболее отдалённой точке $B(6; 0)$ (Следует также учесть, что коэффициент при $(x_1-2)^2$ равен 3, а при $(x_2-3)^2$ коэффициент равен 2). Окружность с центром точке $O_1(2; 3)$ радиуса O_1A содержит вершины O и A области D.

$U_{max}(B)=U_{max}(6; 0)=3(6-2)^2+2(0-3)^2=66;\ x_1=6;\ x_2=0.$

Ответ: $U_{max}(6; 0)=66;\ x_1=6;\ x_2=0.$

Пример7.

Определить критические точки, тип экстремума в них и величину дохода, если известно, что связь между доходом R предприятия за отчетный период, издержками C и прибылью P выражается функциональной зависимостью.

$$R(C, P) = 7C^3 + 6P^3 - 3C^2P + 3CP^2 - 2C.$$

Необходимые условия экстремума: $\begin{cases} R'_C(C,P)=0, \\ R'_p(C,P)=0. \end{cases}$

Находим

$$\begin{cases} R'_C = 21C^2 - 6CP + 3P^2 - 2 \\ R'_P = 18P^2 - 3C^2 + 6CP \end{cases} \Rightarrow \begin{cases} 21C^2 - 6CP + 3P^2 = 2 \\ 18P^2 - 3C^2 + 6CP = 0 \end{cases} \qquad (4.4.3)$$

Сложим почленно уравнения системы (4.4.3):

$$21P^2 + 18C^2 = 2 \qquad (4.4.4)$$

Преобразуем второе уравнение системы (4.4.4):

$6P^2 - C^2 + 2CP = 0$ или $6P^2 - (C - P)^2 + P^2 = 0$, откуда $P = \dfrac{C}{1 \pm \sqrt{7}}$

1) $P = \dfrac{C}{1-\sqrt{7}} = -0,608C$, 2) $P = \dfrac{C}{1+\sqrt{7}} = 0,274C.$

Выражения P через C подставим в (4.4.4):

1) $(21(-0,608)^2 + 18)C^2 = 2,\ C_1 = 0,279$

2) $21(0,274)^2 + 18)C^2 = 2,\ C_2 = 0,320$

В силу экономического смысла $C > 0$.

Отсюда $P_1 = -0{,}608 \cdot 0{,}279 = -0{,}17$, $P_2 = 0{,}32 \cdot 0{,}274 = 0{,}088$.

Критические точки: $M_1\left(\overset{C}{0{,}279};\overset{P}{-0{,}17}\right), M_2\left(\overset{C}{0{,}32};\overset{P}{0{,}088}\right)$.

Используя достаточные условия экстремума, находим:

$$A = R''_{CC}\big|_{M_i} = \left(42C - 6P\right)\big|_{M_i}, \quad C = R''_{PP}\big|_{M_i} = \left(36P + 6C\right)\big|_{M_i} \quad B = R''_{PC}\big|_{M_i} = \left(-6C + 6P\right)\big|_{M_i}.$$

В каждой из точек M_i ($i = 1{,}2$) находим значение выражения $\Delta = AC - B^2$.

$$\Delta\big|_{M_1} = \left(AC - B^2\right)\big|_{M_1} = \left(42\cdot 0{,}279 + 6\cdot 0{,}17\right)\left(36(-0{,}17) + 6\cdot 0{,}279\right) -$$

$-\left(-6\cdot 0{,}279 + 6(-0{,}17)\right)^2 = -63{,}89 < 0 \Rightarrow$ в точке M_1 экстремума нет.

$$\Delta\big|_{M2} = \left(AC - B^2\right)\big|_{M2} = \left(42\cdot 0{,}32 - 6\cdot 0{,}088\right)\left(36\cdot 0{,}088 + 6\cdot 0{,}32\right) -$$

$$-\left(-6\cdot 0{,}32 + 6\cdot 0{,}088\right)^2 = 14{,}176 > 0 \Rightarrow$$

в точке M_2 экстремум есть. Причем, поскольку $A = 12{,}912 > 0$, в точке M_2 функция $R(C, P)$ имеет минимум.

$R_{\min}(0{,}32;\ 0{,}088) = 7(0{,}32)^3 + 6(0{,}088)^3 - 3\cdot 0{,}32^2\cdot 0{,}088 + 3\cdot 0{,}32\cdot(0{,}088)^2 - 2\cdot 0{,}32 = -0{,}426$.

4.4.3. Задачи на условный экстремум с использованием функции Лагранжа

Пример 8.

Имеются два актива со случайными эффективностями R_1 и R_2. Возможные значения этих эффективностей и их вероятности приведены в таблице. Функция полезности инвестора имеет вид: $U(R) = 1{,}2R + 0{,}1R^2$.

Необходимо найти оптимальные пропорции x_1, x_2 при условии, что $x_1 + x_2 = 1$, составного актива по критерию ожидаемой полезности.

Вероятность (P)	0,2	0,8
R_1	6%	1,25%
R_2	1%	2,75%

Математические ожидания эффективностей:

$$\bar{R}_1 = \sum_{i=1}^{2} R_{1i}P_i = 6\cdot 0{,}2 + 1{,}25\cdot 0{,}8 = 2{,}2,$$

$$\bar{R}_2 = \sum_{i=1}^{2} R_{2i}P_i = 1\cdot 0{,}2 + 2{,}75\cdot 0{,}8 = 2{,}4.$$

Пусть $R = x_1R_1 + x_2R_2$ — доходность инвестора.

Определим x_1 и x_2 такие, при которых мат. ожидание $M(U(R)) \to max$.

Имеем $\bar{R} = x_1\bar{R}_1 + x_2\bar{R}_2$ — ожидаемая доходность, где $x_1 + x_2 = 1$.

Функция полезности инвестора

$$U(R) = 1{,}2R + 0{,}1R^2 = 1{,}2\,(x_1R_1 + x_2R_2) + 0{,}1(x_1R_1 + x_2R_2)^2.$$

Математические ожидания функции полезности

$$M\big(U(R)\big) = 1{,}2\left(x_1\bar{R}_1 + x_2\bar{R}_2\right) + 0{,}1M\left(x_1R_1 + x_2R_2\right)^2 =$$

$$= 1{,}2\left(x_1\bar{R}_1 + x_2\bar{R}_2\right) + 0{,}1\Big(x_1^2D\left(R_1\right) + x_2^2D\left(R_2\right) + \left(x_1\bar{R}_1 + x_2\bar{R}_2\right)^2\Big),$$

где $D(R_1)$ и $D(R_2)$ — дисперсии R_1 и R_2 соответственно.

Используем функцию Лагранжа $L(x_1, x_2, \lambda) = M(U(R)) + \lambda(x_1 + x_2 - 1)$

Необходимые условия экстремума $\begin{cases} \dfrac{\partial L}{\partial x_1} = 0 \\ \dfrac{\partial L}{\partial x_2} = 0 \\ \dfrac{\partial L}{\partial \lambda} = 0 \end{cases}$

Находим частные производные:

$$\frac{\partial L}{\partial x_1} = 1{,}2\bar{R}_1 + 0{,}1\Big(2x_1D\left(R_1\right) + 2\left(x_1\bar{R}_1 + x_2\bar{R}_2\right)\bar{R}_1\Big) + \lambda$$

$$\frac{\partial L}{\partial x_2} = 1{,}2\bar{R}_2 + 0{,}1\Big(2x_2D\left(R_2\right) + 2\left(x_1\bar{R}_1 + x_2\bar{R}_2\right)\bar{R}_2\Big) + \lambda$$

$$\frac{\partial L}{\partial \lambda} = x_1 + x_2 - 1$$

Приравняв частные производные нулю, исключим параметр λ и запишем систему уравнений:

$$\begin{cases} 1{,}2\bar{R}_1 + 0{,}1\Big(2x_1D\left(R_1\right) + 2\left(x_1\bar{R}_1 + x_2\bar{R}_2\right)\bar{R}_1\Big) = \\ = 1{,}2\bar{R}_2 + 0{,}1\Big(2x_2D\left(R_2\right) + 2\left(x_1\bar{R}_1 + x_2\bar{R}_2\right)\bar{R}_2\Big) \\ x_1 + x_2 = 1 \end{cases}$$

Приняв $x_2=1-x_1$, преобразуем первое уравнение системы.

$$6R_1 + x_1D\left(R_1\right) + x_1\bar{R}_1^2 + \left(1 - x_1\right)\bar{R}_1\bar{R}_2 =$$

$$= 6R_2 - \left(1 - x_1\right)D\left(R_2\right) + x_1\bar{R}_1\bar{R}_2 + \left(1 - x_1\right)\bar{R}_2^2 =,$$

Отсюда $$x_1 = \frac{6\left(R_2 - R_1\right) - \bar{R}_1\bar{R}_2 - D\left(R_2\right) + \bar{R}_2^2}{D\left(R_1\right) + D\left(R_2\right) + \left(\bar{R}_2 - \bar{R}_1\right)^2}$$

Вычислим $D(R_1)$ и $D(R_2)$:

$$D(R_1) = \sum_{i=1}^{2} R_{1i}^2P_i - \bar{R}_1^2 = 6^2 \cdot 0{,}2 + 1{,}25^2 \cdot 0{,}8 - 2{,}2^2 = 3{,}61$$

$$D(R_2)=\sum_{i=1}^{2}R_{2i}^2P_i-\overline{R}_2^2=1^2\cdot 0,2+2,75^2\cdot 0,8-2,4^2=0,49$$

Находим

$$x_1=\frac{6(2,4-2,2)-2,2\cdot 2,4-0,49+2,4^2}{3,61+0,49-(2,4-2,2)^2}=\frac{1,19}{4,14}=0,287,$$

$$x_2=1-x_1=1-0,287=0,713.$$

Поскольку $\overline{R}_2>\overline{R}_1$, т.е. $D\overline{R}_2>D\overline{R}_1$ степень риска для R_2 меньше, чем для R_1, получено, что $x_2>x_1$.

Ответ: $x_1 = 0,287; x_2 = 0,426.$

4.5. ФУНКЦИИ НЕСКОЛЬКИХ ПЕРЕМЕННЫХ В ЭКОНОМИЧЕСКОЙ ТЕОРИИ

Рассмотрим некоторые приложения функций нескольких переменных в экономической теории.

Функция полезности – одно из базовых понятий экономической теории. Многомерный ее аналог – это функция $z=f(x_1,\ldots,x_n)$, выражающая полезность от n приобретенных товаров. Чаще всего встречаются следующие ее виды:

а) $z=\sum_{i=1}^{n}a_i\ln(x_i-c_i)$, *где* $a_i>0,\ x_i>c_i\geq 0$ – *логарифмическая функция*;

б) $z=\sum_{i=1}^{n}\frac{a_i}{1-b_i}(x_i-c_i)^{1-b_i}$. *Здесь* $a_i>0,\ 0<b_i<1;\ x_i>c_i\geq 0.$ Такая функция называется функцией *постоянной эластичности*.

Также на случай n переменных обобщается понятие *производственной функции*, выражающей результат производственной деятельности от обусловивших его факторов $x_1,\ x_2,\ldots,x_n$. Приведем здесь наиболее часто встречающиеся виды производственных функций (z – величина обществен-ного продукта, x_1 – затраты труда, x_2 – объем производственных фондов), полагая для простоты $n=2$:

а) функция Кобба-Дугласа $z=b_0x_1^{b_1}x_2^{b_2}$;

б) функция с постоянной эластичностью замещения: $z=e_0\left[e_1x_1^{-\beta}+e_2x_2^{-\beta}\right]^{-\frac{h}{\beta}}$.

Значительная часть экономических процессов иллюстрируется на рисунках, изображающих линии уровня функции двух переменных $z=f(x,y)$. Например, линии уровня производственной функции называются *изоквантами*.

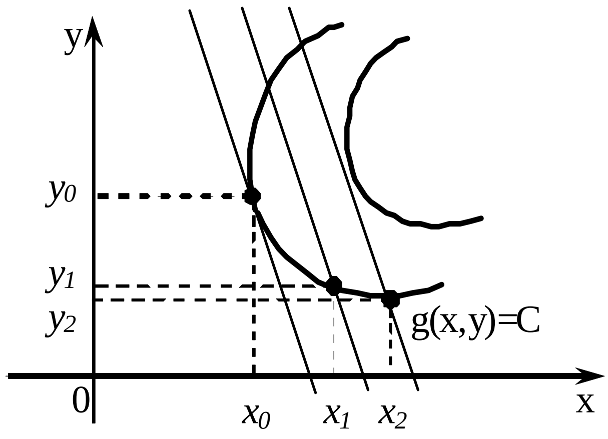


Рис. 4.5

Пусть рассматривается функция выпуска $z=f(x,y)$, где и x, y – два различных фактора произ-водства, а $f(x,y)$ – максимально возможный выпуск продукции, который позволяют значения факторов x и y. Все значения факторов (переменных) должны принадлежать так называемой экономической области, которая характеризуется тем, что высекаемые ею части изоквант представляют собой графики убывающих функций, т.е. увеличение количества одного фактора позволяет уменьшить количество другого, не меняя размера выпуска. Иными словами, экономическая область – это множество значений факторов, допускающих замещение одного из них другим.

Изокванты позволяют геометрически иллюстрировать решение задачи об оптимальном распределении ресурсов. Пусть $z=g(x,y)$ – функция издержек, характеризующая затраты, необходимые для обеспечения значений ресурсов x и y (часто можно считать, что функция издержек линейная: $g(x,y)=p_x x+p_y y$, где p_x и p_y – "цены" факторов x и y). Линии уровня этой функции – прямые, изображены на рис.4.5. Комбинация линий уровня функции $f(x)$ и $g(x)$ позволяют делать выводы о предпочтительности того или иного значения факторов x и y. Очевидно, например, что пара значений (x_1, y_1) более предпочтительна, чем пара (x_2, y_2), так как обеспечивает тот же выпуск, но с меньшими затратами. Оптимальными же значениями факторов будут значения (x_0, y_0) – координаты точки касания линии уровня функции выпуска и функции издержек.

Линии уровня функции полезности (они называются кривыми безразличия) также позволяют рассматривать вопросы замещения одного товара другим и иллюстрировать решение задачи об оптимальном потреблении (рис. 4.6).

Линии уровня затрат на приобретение товаров x, y изображены на рис. 4.6 пунктиром. Оптимальное потребление обеспечивается значением (x_0, y_0) – координатами точки касания кривой безразличия и линии уровня затрат. В этой точке заданная полезность достигается наиболее эконо-мичным образом.

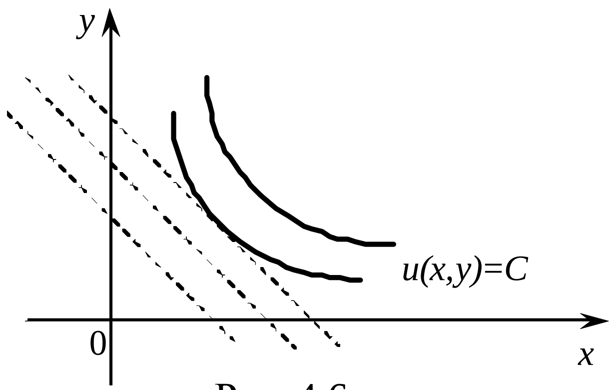


Рис. 4.6

Понятие частной производной также находит применение в экономической теории. Ранее было введено понятие эластичности функции одной переменной $E_x(y)$. Аналогично можно ввести понятие частной эластичности функции нескольких переменных. $z=(x_1,x_2,\ldots,x_n)$ относительно переменной x_i:

$$E_{x_i}(z) = \lim_{\Delta x_i \to 0} \left(\frac{\Delta x_i z}{z} : \frac{\Delta x_i}{x_i} \right) = \frac{x_i}{z} \cdot z'_{x_i}.$$

Так, например, в производственной функции Кобба-Дугласа $z = b_0 x^{b_1} y^{b_2}$, как нетрудно убедиться, $E_x(z)=b_1$, $E_y(z)=b_2$, т.е. показатели b_1 и b_2 приближенно показывают, на сколько процентов изменится выпуск продукции при изменении только затрат труда x или только объема производственных фондов y на 1%.

Рассмотрим теперь частные производные u'_x, u'_y — функции полезности. Они называются предельными (маргинальными) полезностями Mu_x, Mu_y. Если измерять количество товара в стоимостном выражении, то предельные полезности можно рассматривать как функции спроса на сопутствующий товар. Найдем предельные полезности для функции постоянной эластичности

$$u(x,y) = \frac{a_1}{1-b_1} x^{1-b_1} + \frac{a_2}{1-b_2} y^{1-b_2}.$$

Имеем $Mu_x = a_1 x^{-b_1}$, $Mu_y = a_2 x^{-b_2}$, т.е. функции спроса с ростом стоимости каждого товара являются убывающими, а параметры b_1 и b_2 представляют частные эластичности спроса на эти товары.

Пример 9. Для выпуска некоторого товара определена производственная функция $f(x,y)=20x+10y-2y^2+4x^2+3xy$, где x, y – факторы производства. Определить: а) закон изменения производственной функции; б) эластичность функции по каждому фактору; в) коэффициент эластичности по факторам при $x=1, y=1$.

а) Чтобы определить изменение производственной функции по факторам x и y соответственно, необходимо найти $\frac{\partial f}{\partial x}$ и $\frac{\partial f}{\partial y}$:

$$\frac{\partial f}{\partial x} = 20 + 8x + 3y; \qquad \frac{\partial f}{\partial y} = 10 - 4y + 3x.$$

б) По определению эластичность функции по каждому из факторов такова:

$$E_x(z) = \frac{x}{z}\frac{\partial z}{\partial x} = \frac{x}{z}(20 + 8x + 3y); \quad E_y(z) = \frac{y}{z}\frac{\partial z}{\partial y} = \frac{y}{z}(10 - 4y + 3x),$$

где $z=20x+10y-2y^2+4x^2+3xy$.

в) Вычислим коэффициенты эластичности при $x=1$, $y=1$. Учтем, что $z(1,1)=20+10-2+4+3=35$:

$$E_x(z) = \frac{20+8+3}{35} = \frac{31}{35} \approx 0{,}89; \quad E_y(z) = \frac{10-4+3}{35} = \frac{9}{35} \approx 0{,}26.$$

Таким образом, с увеличением фактора x на 1% произойдет относительное увеличение заданной производственной функции приближенно на 0,89% (при условии стабильности фактора y).

При увеличении фактора на 1% и неизменности фактора x производственная функция увеличится приближенно на 0,26%. Значит, наибольшее влияние на производственную функцию $z=f(x,y)$ оказывает фактор x.

Отметим, что отрицательное значение коэффициента эластичности показывает уменьшение производственной функции при увеличении соответствующего фактора. Например, если $E_x(z)=-0{,}08$ и $z=f(x,y)$ – функция выпуска продукции, то увеличение фактора x на 1% приводит к снижению выпуска продукции на 0,08%.

Пример 10. Фирма производит два вида товаров G_1 и G_2 и продает их по цене 1000 ден.ед. и 800 ден.ед. соответственно. Объемы выпуска товаров – Q_1 и Q_2. Функция затрат имеет вид: $C = 2Q_1^2 + 2Q_1Q_2 + Q_2^2$. Требуется найти такие значения Q_1 и Q_2, при которых прибыль, получаемая фирмой, максимальна, и найти эту прибыль.

Суммарный доход от продажи товаров G_1 и G_2: $R=1000Q_1+800Q_2$.

Прибыль $П$ представляет собой разницу между доходом R и затратами C, поэтому $П = R - C = (1000Q_1 + 800Q_2) - (2Q_1^2 + 2Q_1Q_2 + Q_2^2)$, или

$$П(Q_1,Q_2) = 1000Q_1 + 800Q_2 - 2Q_1^2 - 2Q_1Q_2 - Q_2^2.$$

Это и есть функция, максимум которой следует найти.

Для нахождения стационарных точек, определяем частные производные первого порядка от функции $П(Q_1,Q_2)$ и приравниваем их к нулю:

$$П'_{Q_1}(Q_1,Q_2) = 1000 - 4Q_1 - 2Q_2;\ П'_{Q_2}(Q_1,Q_2) = 800 - 2Q_1 - 2Q_2;$$

$$\begin{cases} 1000 - 4Q_1 - 2Q_2 = 0; \\ 800 - 2Q_1 - 2Q_2 = 0. \end{cases}$$

Решение системы: $Q_1=100$, $Q_2=300$, стационарная точка $M_0(100;300)$.

Находим частные производные второго порядка:

$$A = \frac{\partial^2 П}{\partial Q_1^2}\Big|_{M_0} = -4;\ B = \frac{\partial^2 П}{\partial Q_1 \partial Q_2}\Big|_{M_0} = -2;\ C = \frac{\partial^2 П}{\partial^2 Q_2^2}\Big|_{M_0} = -2,$$

$$A < 0,\ \Delta = AC - B^2 = -4 \cdot (-2) - 4 = 4 > 0.$$

Значит, точка $M_0(100;300)$ является точкой максимума.

Максимальная прибыль достигается при объемах производства $Q_1=100$, $Q_2=300$. Найдем сумму максимальной прибыли:

$$П(100;300) = 1000 \cdot 100 + 800 \cdot 300 - 2 \cdot 100^2 - 2 \cdot 100 \cdot 300^2 = 170\,000\,ден.ед.$$

Понятие выпуклости функции также играет существенную роль в понимании важнейших экономических законов. Многомерные аналоги примеров, рассмотренных выше, позволяют математически сформулировать законы убывающей доходности и убывающей предельной полезности.

РАЗДЕЛ 5. МЕТОД НАИМЕНЬШИХ КВАДРАТОВ

5.1. ИНТЕРПОЛИРОВАНИЕ

Если задана функция $y(x)$, то это означает, что любому допустимому значению x сопоставлено значение y. Но нередко оказывается, что нахождение этого значения очень трудоемко или невозможно. Например, $y(x)$ может быть определено как решение сложной задачи, в которой x играет роль параметра, при этом можно вычислить небольшую таблицу значений функции, но прямое нахождение функции при большом числе значений аргумента практически невозможно. В этом случае функцию $y(x)$ заменяют приближенной формулой, т.е. подбирают некоторую функцию $\varphi(x)$, принадлежащую некоторому классу, которая при некоторых значениях аргумента принимает те же значения, что и $y(x)$. Полагают $y(x_i)\approx \varphi(x_i)$, где x_i, принадлежащие некоторому отрезку $[a,b]$, называются узлами интерполяции, а $\varphi(x)$ – интерполирующая функция. Геометрически это означает, что нужно найти кривую $y= \varphi(x)$ некоторого типа, проходящую через заданную систему точек $M(x_i, y_i)$ $(i=0,1,2,\ldots,n)$. Близость получают введением в аппроксимирующую функцию свободных параметров $a=\{a_1, a_2,\ldots, a_n\}$ и соответствующим их набором.

Подбор удачного вида функциональной зависимости $\varphi(x,a)$ – это проблема, которая будет рассмотрена ниже.

5.1.1. Линейная интерполяция

Пусть функция $y(x)$ известна только в узлах некоторой сетки x_i , т.е. задана таблицей. Если потребовать, чтобы $\varphi(x,a)$совпадала с табличными значениями в n выбранных узлах сетки, то получим систему

$$\varphi(x_i;\, a_1, a_2,\ldots, a_n)=y(x_i)\equiv y_i, \qquad i=\overline{1,n}. \tag{5.1.1}$$

из которой можно определить параметры a_k. Этот способ подбора параметров называется интерполяцией (точнее, лагранжевой интерполяцией).

Интерполяция называется линейной, если $\varphi(x,a)$ линейно зависит от параметров, т.е. представима в виде так называемого обобщенного многочлена

$$\varphi(x;a_1,a_2,...,a_n)=\sum_{k=1}^{n}a_k\varphi_k(x), \tag{5.1.2}$$

где $\varphi_k(x)$ – линейно-независимы.

Подставляя (5.1.2) в (5.1.1), получим систему линейных уравнений:

$$\sum_{k=1}^{n} a_k \varphi_k(x_i) = y_i, \quad i = \overline{1,n} \tag{5.1.3}$$

Чтобы задача интерполирования всегда имела единственное решение, надо, чтобы при произвольном расположении узлов (лишь бы среди них не было совпадающих) определитель системы (5.1.3) был отличен от нуля:

$$Det\{\varphi_k(x_i)\} = \begin{vmatrix} \varphi_1(x_1) & \varphi_2(x_1) & \dots & \varphi_n(x_1) \\ \varphi_1(x_2) & \varphi_2(x_2) & \dots & \varphi_n(x_2) \\ \dots & \dots & \dots & \dots \\ \varphi_1(x_n) & \varphi_2(x_n) & \dots & \varphi_n(x_n) \end{vmatrix} \neq 0 \quad \text{при } x_i \neq x_j \tag{5.1.4}$$

Система функций, удовлетворяющих требованию (5.1.4), называется чебышевской. Таким образом, при линейной интерполяции надо строить обобщенный многочлен по какой-нибудь чебышевской системе функций.

5.2. СРЕДНЕКВАДРАТИЧНОЕ ПРИБЛИЖЕНИЕ

5.2.1. Сглаживание экспериментальных зависимостей

На практике мы часто сталкиваемся с задачей о сглаживании экспериментальных зависимостей.

Пусть зависимость между двумя переменными x и y выражается в виде таблицы, полученной опытным путем. Это могут быть результаты опыта или наблюдений, статистической обработки материала и т.п.

x	x_1	x_2	…	x_i	…	x_n
y	y_1	y_2	…	y_i	…	y_n

Требуется наилучшим образом сгладить экспериментальную зависимость между переменными x и y, т.е. по возможности точно отразить общую тенденцию зависимости y от x, исключив при этом случайные отклонения, связанные с неизбежными погрешностями измерений или статистических наблюдений. Такую сглаженную зависимость стремятся представить в виде формулы $y=f(x)$.

Формулы, служащие для аналитического представления опытных данных, получили название эмпирических формул.

Задача, нахождения эмпирических формул разбивается на два этапа. На первом этапе нужно установить вид зависимости $y=f(x)$, т.е. решить, является ли она линейной, квадратичной, логарифмической или какой-либо другой. Предположим, например, что результаты экспериментальных исследований нанесены на плоскость (паре чисел (x_i, y_i) соответствует точка с такими же координатами). Разумеется, существует множество кривых, проходящих через эти точки. Для про-движения к цели обычно предполагают, что кривая истинной зависимости – это наиболее "гладкая" кривая, согласованная с эмпирическими данными. Так, в случае, изображенном на рис. 5.1, исследователь несомненно предпочтет кривую I кривой II.

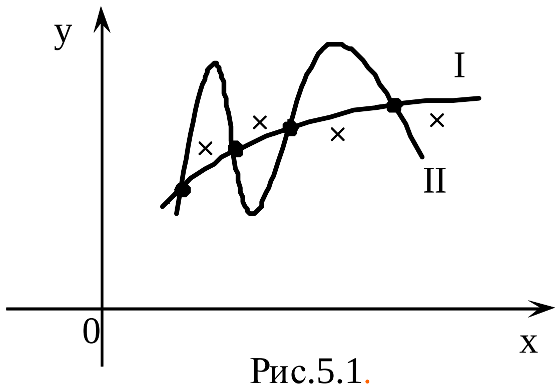


Рис.5.1.

Для проверки правильности вывода проводятся дополнительные исследования, т.е. производится еще ряд одновременных измерений величин x и у. Дополнительные точки наносятся на плоскость. Если они оказываются достаточно близкими к выбранной кривой (на рисунке допол-нительные точки изображены крестиками), то можно считать, что вид кривой установлен. В противном случае кривую надо скорректировать и вновь провести дополнительные измерения.

Кроме того, для выбора функции $y=f(x)$ привлекаются дополнительные соображения, как правило, не математического характера (теоретические предпосылки, опыт предшествующих исследований и т.п.).

Предположим, первый этап завершен – вид функции $y=f(x)$ выбран. Тогда переходят ко второму этапу – определению неизвестных параметров этой функции.

Согласно наиболее распространенному и теоретически обоснованному методу наименьших квадратов в качестве неизвестных параметров функции $f(x)$ выбирают такие значения, чтобы сумма квадратов невязок δ_i, или отклонений "теоретических" значений $f(x_i)$, найденных по эмпирической формуле $y=f(x)$, от соответствующих опытных значений y_i, т.е. $S=\sum_{i=1}^{n}\delta_i^2=\sum_{i=1}^{n}(f(x_i)-y_i)^2$ была минимальной (рис.5.2).

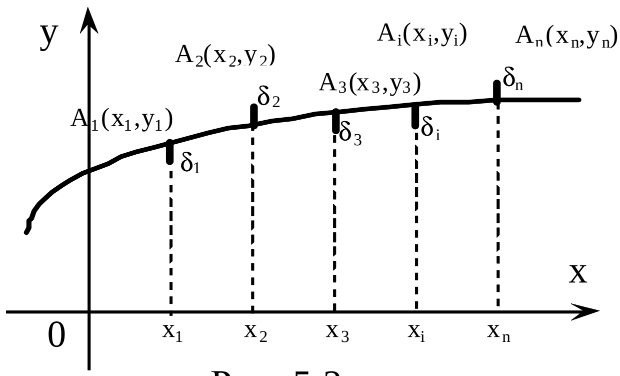


Рис. 5.2

Следует отметить, что в качестве величины отклонения δ эмпирических точек (x_i,y_i) от точек сглаживающей экспериментальную зависимость кривой $y=f(x)$ в принципе можно было взять обычную сумму невязок

$$\sum_{i=1}^{n}\delta_i=\sum_{i=1}^{n}(f(x_i)-y_i)$$

или сумму их абсолютных величин

$$\sum_{i=1}^{n}|\delta_i|=\sum_{i=1}^{n}|(f(x_i)-y_i)|.$$

Но делать это нецелесообразно, так как в первом случае $\sum_{i=1}^{n}\delta_i$ может быть малой или даже равняться нулю при зна-чительном разбросе эмпирических точек, так как положительные откло-нения δ_i компенсируются отрицатель-ными. Во втором случае функция $\sum_{i=1}^{n}|\delta_i|$ лишена этого недостатка, но имеет другой: она не является дифферен-цируемой, что существенно затрудняет решение задачи.

Пусть в качестве функции $y=f(x)$ взята линейная функция $y=ax+b$ и задача сводится к отысканию таких значений параметров a и b, при которых функция $S=\sum_{i=1}^{n}(ax_i+b-y_i)^2$ принимает наименьшее значение. Заметим, что функция $S=S(a;b)$ есть функция двух переменных a и b до тех пор, пока мы не нашли, а затем зафиксировали их "наилучшие" (в смысле метода наименьших квадратов) значения, а x_i, y_i – постоянные числа, найденные экспериментально. Таким образом, для нахождения прямой, наилучшим образом согласованной с опытными данными, достаточно решить систему

$$\begin{cases} S'_x = 0, \\ S'_y = 0, \end{cases} \text{ или } \begin{cases} \sum\limits_{i=1}^{n} 2(ax_i + b - y_i)x_i = 0, \\ \sum\limits_{i=1}^{n} 2(ax_i + b - y_i) = 0. \end{cases} \qquad (5.2.1)$$

После алгебраических преобразований эта система принимает вид:

$$\begin{cases} \left(\sum\limits_{i=1}^{n} x_i^2\right)a + \left(\sum\limits_{i=1}^{n} x_i\right)b = \sum\limits_{i=1}^{n} x_i y_i, \\ \left(\sum\limits_{i=1}^{n} x_i\right)a \ + \ nb \quad = \sum\limits_{i=1}^{n} y_1. \end{cases} \qquad (5.2.2)$$

Эта система называется системой нормальных уравнений и имеет единственное решение, так как ее определитель

$$|A| = \begin{vmatrix} \sum\limits_{i=1}^{n} x_i^2 & \sum\limits_{i=1}^{n} x_i \\ \sum\limits_{i=1}^{n} x_i & n \end{vmatrix} = n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2 \neq 0 \text{ (а точнее } |A| > 0\text{)}.$$

Убедимся, что найденные значения a , b дают минимум функции $S=S(a;b)$. Найдем вторые частные производные

$$S''_{aa} = 2\sum_{i=1}^{n} x_i^2 = A;\ S''_{ab} = 2\sum_{i=1}^{n} x_i = B;\ S''_{bb} = 2n = C.$$

Выражение $\Delta = AB - C^2 = 4\left(n\sum\limits_{i=1}^{n} x_i^2 - \left(\sum\limits_{i=1}^{n} x_i\right)^2\right) > 0$ в силу изложенного выше и $A = 2\sum\limits_{i=1}^{n} x_i^2 > 0$, следовательно, согласно достаточному условию, функция имеет единственную точку минимума, определяемую из системы нормальных уравнений. Заметим, что в этой точке функция $S=S(a;b)$ имеет не просто локальный минимум, но наименьшее значение (глобальный минимум).

Пример 1. Имеются следующие данные о цене на нефть *x* (ден. ед.) и индекс нефтяных компаний *y (у.ед.)*.

x	17,28	17,05	18,30	18,80	19,20	18,50
y	537	534	550	555	560	552

Предполагая, что между переменными x и y существует линейная зависимость, найти эмпирическую формулу вида $y=ax+b$, используя метод наименьших квадратов. Промежуточные вычисления оформим в виде таблицы.

x_i	y_i	$x_i y_i$	x_i^2
17,28	537	9279,36	298,5984
17,05	534	9104,70	290,7025
18,30	550	10065,00	334,8900
18,80	555	10434,00	353,4400
19,20	560	10752,00	368,6400
18,50	552	10212,00	342,2500
∑110,13	3288	59847,06	1988,5200

Система нормальных уравнений имеет вид

$$\begin{cases} 1988,52a + 110,13b = 59847,06; \\ 110,13a + 6b = 3288. \end{cases}$$

Ее решение $a=15,317$, $b=266,86$ дает искомую зависимость: $y=15,317x+266,86$. Таким образом, с увеличением цены нефти на 1 ден. ед. индекс акций нефтяных компаний в среднем растет на 15,32 ед.

5.2.2. Наилучшее приближение

Интерполяция позволяет легко аппроксимировать функцию $y(x)$. Однако точность такой аппроксимации гарантирована лишь в небольшом интервале порядка нескольких шагов сетки. Для другого интервала приходится заново вычислять коэффициенты интерполяционной формулы. Нам желательно иметь единую приближенную формулу $y(x)\approx \varphi(x)$, пригодную для произвольного отрезка.

При интерполировании мы приравниваем значения $y(x)$ и $\varphi(x)$ в узлах. Если $y(x_i)$ определены неточно, – например, из эксперимента, – то точное приравнивание неразумно. Поэтому нередко целесообразнее приближать не по точкам, а в среднем, т.е. по норме L_p[1].

Пусть заданы функция $y(x)$ и множество функций $\varphi(x)$, принадлежащие линейному нормированному пространству функций. Нас интересует две задачи. Первая – аппроксимация с заданной точностью по заданному значению ε найти такую $\varphi(x)$, чтобы выполнялось неравенство

[1] Введена метрика

$\|y(x)\text{-}\varphi(x)\|\leq\varepsilon$. Второе – нахождение наилучшего приближения, т.е. функции $\varphi^*(x)$, удовлетворяющей соотношению

$$\left\|y(x)-\varphi^*(x)\right\| = \inf\left\|y(x)-\varphi(x)\right\| = \nu \tag{5.2.3}$$

Возникает вопрос, существует ли наилучшее приближение и единственно ли оно (для данных функций и множества)?

Доказано, что в любом линейном нормированном пространстве при линейной аппроксимации

$$\varphi(x) = \sum_{k=1}^{n} a_k\varphi_k(x)\,, \tag{5.2.4}$$

где $\varphi_k(x)$ – линейно-независимы, наилучшее приближение существует, хотя не во всяком линейном пространстве оно единственно.

На практике используются пространства L_2 и C[2].

5.2.3. Линейная аппроксимация

Рассмотрим гильбертово пространство $L_2(\rho)$ действительных функций, интегрируемых с квадратом с весом $\rho(x)>0$ на $[a,b]$, норма в нем равна $\|f\|_{L_2} = \sqrt{(f,f)}$, где скалярное произведение определено следующим образом:

$$(f,\varphi) = \int_a^b \rho(x)f(x)\varphi(x)dx\,.$$

Выберем в качестве аппроксимирующей функции линейную комбинацию (5.2.4). Подставив ее в условие наилучшего приближения (5.2.3), получим

$$\left\|y-\varphi\right\|_{L_2} = (y,y) - 2\sum_{k=1}^{n} a_k(y,\varphi_k) + \sum_{k,m=1}^{n} a_k a_m(\varphi_k,\varphi_m) = \min \tag{5.2.5}$$

Используя необходимые условия существования экстремума функции многих переменных, приравняем к нулю частные производные по коэффициентам и получим систему линейных уравнений

$$\sum_{m=1}^{n}(\varphi_k,\varphi_m)a_m = (y,\varphi_k), \quad k=\overline{1,n} \tag{5.2.6}$$

Её определитель (определитель Грамма) функций $\varphi_k(x)$ отличен от нуля, так как функции $\varphi_k(x)$ – линейно независимы. Следовательно, наилучшее среднеквадратичное приближение существует и единственно. Для его нахождения надо решить систему линейных алгебраических уравнений.

2

Если вещественные функции заданы таблично, т.е. на конечном множестве точек, то их скалярное произведение определяется формулой

$$(f\varphi) = \sum_{i=1}^{N} \rho_i f(x_i)\varphi(x_i), \quad \rho_i > 0 \ , \tag{5.2.7}$$

где N – полное число узлов таблицы. Тогда условие наилучшего среднеквадратичного приближения (5.2.3) примет вид

$$\delta_\varphi^2 \sum_{i=1}^{N} \rho_i \equiv \sum_{i-1}^{N} \rho_i \left[y(x_i) - \varphi(x_i) \right]^2 = \min \qquad (5.2.8).$$

Выберем линейную аппроксимацию $\varphi(x) = \sum_{k=1}^{n} a_k \varphi_k(x)$ с числом членов $n \le N$. Тогда коэффициенты аппроксимации находятся из уравнений (5.2.6), где скалярное произведение надо брать по формуле (5.2.7); эти уравнения можно получить и непосредственно, подставляя обобщенный многочлен в (5.2.8) и приравнивая нулю производные по коэффициентам. Описанный метод аппроксимации называется ***методом наименьших квадратов.***

Метод наименьших квадратов широко используют для обработки экспериментальных кривых, точки которых измерены с заметной погрешностью ε. В этом случае весу ρ_i придают смысл точности измерения в данной точке: чем выше точность, тем большее значение веса приписывают точке (обычно выбирают $\rho_i = \frac{1}{\varepsilon_i^2}$). Аппроксимирующая кривая будет проходить ближе к точкам с большим весом.

5.3. МНК – ОСНОВНОЙ МЕТОД МАТЕМАТИЧЕСКОГО МОДЕЛИРОВАНИЯ В ЭКОНОМИКЕ

Экономическая теория – это мощный инструмент исследования, анали-за, объяснения и прогнозирования динамики экономической реальности. Современная экономика с разнообразием ее подходов и способов наблюдения, методов обработки информации и моделирования экономических систем стала в большей мере междисциплинарным творением. Особое место в экономике занимает эконометрика, в буквальном переводе – «измерение экономики».

«Эконометрика – это наука, которая изучает количественные законно-мерности и взаимосвязи экономических объектов с помощью математико-

статистических методов и моделей и их анализа для выработки конкретных рекомендаций».

В экономике в большинстве случаев между переменными величинами существуют зависимости, когда каждому значению одной переменной соответствует не какое-то определённое, а множество возможных значений другой переменной. Иначе говоря, каждому значению одной переменной соответствует *определённое (условное) распределение другой переменной*.

Такая зависимость получила название статистической (вероятностной).

В силу неоднозначности статистической зависимости между X и $У$ представляет интерес *усреднённая* по X схема зависимости, т.е. законно-мерность в изменении условного математического ожидания $M_x(У)$, которая называется корреляционной зависимостью.

Корреляционная зависимость может быть представлена в виде:
$M_x(У)=\varphi(x)$ или $M_y(X)=\psi(y)$, где $\varphi(x)\neq const$, $\psi(y\neq const$.

Эти уравнения называются модельными уравнениями регрессии (или просто *уравнениями регрессии)* соответственно $У$ по X или X по $У$, а их графики *линиями регрессии*.

Для отыскания модельных уравнений регрессии, вообще говоря, необходимо знать *закон распределения двумерной случайной величины* $(X,У)$. На практике исследователь, как правило, располагает лишь выборкой пар значений (x_i, y_i) ограниченного объёма. В этом случае речь может идти об оценке (приближённом выражении) по выборке функции регрессии. Такой наилучшей в смысле *метода наименьших квадратов* оценкой является выборочная линия (кривая) регрессии $У$ по X $y_x=\varphi(x,a_0,a_1,\ldots,a_k)$, где y_x– *условная (групповая) средняя переменной* $У$ при фиксированном значении переменной $X=x$; $a_0,a_1,\ldots,a_k$–параметры кривой.

5.3.1. Основные определения. Примеры эконометрических моделей

Методы изучения реальных процессов и систем – натурный эксперимент и моделирование. Ограниченные возможности натурного эксперимента очевидны. В связи с этим основным средством исследования реального мира является моделирование. Модель – это абстрактный объект. На практике используют два вида математических моделей.

Макромодели − ‘‘черный ящик’’, устанавливающий связь между входом $X(t)$ и выходом $Y(t)$ в виде соотношения: $X(t)\rightarrow Y(t)$

Микромодели – известен механизм преобразования входа в выход.

$X(t)$ – вектор-функция входных параметров.

$Y(t)$ – вектор- функция выходных параметров.

$S(t)$ – вектор, описывающий множество состояний систем.

$G(t)$ – оператор, переводящий систему из одного состояния в другое в зависимости от входа $X(t)$ и предыдущего состояния. Например, $S(t+1)=G(S(t), X(t))$ или $S(t)=G(S(t); X(t))$.

$H(t)$–оператор, определяющий выход системы в зависимости от состояния системы и входа. При этом, $Y(t)= H(S(t), X(t))$. Приведенные соотношения образуют микромодель.

Пример 1. Производственная система

Зафиксируем для простоты t, т.е., примем: $X(t)=X; Y(t)= y; S(t) = S$.

Пусть $B= (b_1,b_2,\dots b_m)^T$ – вектор расходуемых ресурсов (сырье, электроэнергия и т.п.).

$X=(x_1,x_2,\dots x_n)^T$ – план производств.

$S=(s_1,s_2,\dots s_n)^T$ – вектор номенклатуры, выпускаемой продукции.

$$G=\begin{pmatrix} g_{11} & g_{12} & \cdots & g_{1n} \\ g_{21} & g_{22} & \cdots & g_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ g_{m1} & g_{m2} & \cdots & g_{mn} \end{pmatrix}$$ – матрица связи между затратами ресурсов и номенклатурой, т.е. g_{ij} – расход i - го ресурса на изготовление единицы номенклатуры j-го вида.

При этом $GX=B$, или

$g_{i1}x_1+ g_{i2}x_2+\dots+ g_{in}x_n =b_i \quad \overline{i= 1,m}$.

$H = (h_1, h_2, \dots h_n)^T$– вектор задающий выход в зависимости от состояния.

h_j – прибыль от реализации единицы номенклатуры j- го вида.

$У=H^T \cdot X=h_1x_1+ h_2x_2+\dots+ h_nx_n$ – суммарная прибыль.

Пример 2. Предложение и спрос на конкурентном рынке

q_c – спрос

q_n – предложение

p – цена продукта

$q_{c=} f_c(P)$ – функция спроса;

$q_{n=} f_n(P)$ – функция предложения;

$f_c(P) = f_n(P)$ – равновесие

Равновесие на рынке устанавливается в результате некоторого динамического процесса (паутинная модель).

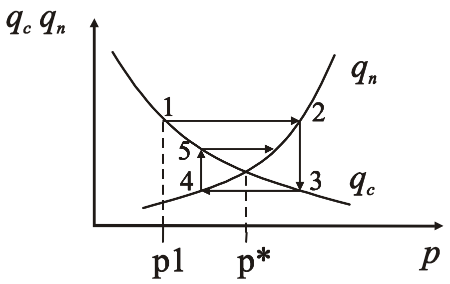


Рис. 5.3

Производитель предоставляет товар на рынок по цене P_1. Спрос по этой цене велик (соотв. точке 1). Большой спрос инициирует повышение цены производителем (точка 2). Это сразу же уменьшает спрос (3), что заставит уменьшаться цену (4). Это приведет к увеличению спроса (5) и т. д.

Пример 3. Модель Кейса

Капиталовложения осуществляют в два сектора (производство средств производства и производство предметов потребления).

Увеличение средств производства приводит к увеличению доходов лиц, участвующих в производстве. Новые доходы порождают увеличение расходов на предметы потребления, что стимулирует их производство.

Введем:

P – общий расход на производство продукции.

C – расход на производство предметов потребления.

I – расход на производство средств производства.

R – доход от производства.

$$\text{Тогда } P = C + I, \tag{5.3.1}$$

$$\text{где: } C = F(R) \tag{5.3.2}$$

$$R = G(P) \tag{5.3.3}$$

$F(R)$ – функция связывающая доход и производство, предмет и потребление.

$G(P)$ – функция, определяющая доход от затрат на производство.

На производство средств производства осуществляется капиталовложение I, тогда получим уравнение,

$$R = G(C+I) = G(F(R) + I), \tag{5.3.4}$$

которое определяет уровень дохода от величины капиталовложения.

Например, если $C=F(R) = \alpha R+\beta; \quad \alpha \in [0;1]$

$$R=G(P)= \gamma P+ \delta; \quad \gamma \in [0;1]$$

то уравнение (5.3.4) принимает вид:

$$R = \gamma\,[\,(\,\alpha R + \beta\,) + I\,] + \delta = \alpha\gamma R + \gamma I + \gamma\beta + \delta$$

Отсюда:

$$R = \frac{\gamma I}{1-\alpha\gamma} + \frac{\gamma\beta + \delta}{1-\alpha\gamma} \qquad (5.3.5)$$

Из последнего соотношения следует, что уровень дохода растет по мере приближения параметров α и γ к единице.

5.3.2. Эндогенные и экзогенные переменные.

Все переменные модели делятся на две категории: эндогенные и экзогенные.

Эндогенные переменные являются объектами объяснения.

Экзогенные переменные объясняют эндогенную, они входят в модель, как независимые переменные. Таким образом, модель отражает определение эндогенных переменных через экзогенные, или, что то же самое, определяет характер влияния экзогенных переменных на эндогенную. Например, в модели Кейса три эндогенные переменные: P, R, C, а I – экзогенная переменная .

5.4. ОПРЕДЕЛЕНИЕ ЗАВИСИМОСТЕЙ МЕЖДУ ЭКОНОМИЧЕСКИМИ ПОКАЗАТЕЛЯМИ

5.4.1. Модели математических зависимостей

Пусть эндогенная переменная y неизвестным образом зависит от набора экзогенных переменных $x_1, x_2, \dots x_n$. Введем неизвестную функцию $y=f(x_1, x_2, \dots x_n)$. Вид этой функции может быть установлен, если вскрыт и известен механизм воздействия экзогенных переменных $X= (x_1, x_2, \dots, x_n)$ на эндогенную переменную y. При этом для описания зависимости между y и $x_1, x_2, \dots, x_n$ могут быть использованы различные соотношения. Рассмотрим некоторые из них применительно к частному случаю, когда экзогенная переменная является единственной.

1. Линейная однофакторная регрессия.

$$y=a_0+a_1 \qquad (5.4.1)$$

2. Нелинейная однофакторная регрессия.

$$y=a_0+a_1x+ a_2x^2+\dots +a_nx^n \qquad (5.4.2)$$

3. Показательная зависимость с линейным показателем.

$$y=d_0b^x \qquad (5.4.3)$$

Частный случай: $y{=}a_0e^x$– экспоненциальная зависимость (5.4.4)

4. Модифицированная показательная зависимость.

$$y{=}a_0{+}a_1b^x \qquad (5.4.5)$$

На рис. 5.4 приведены 4 варианта кривой в зависимости от значений a_1 и b.

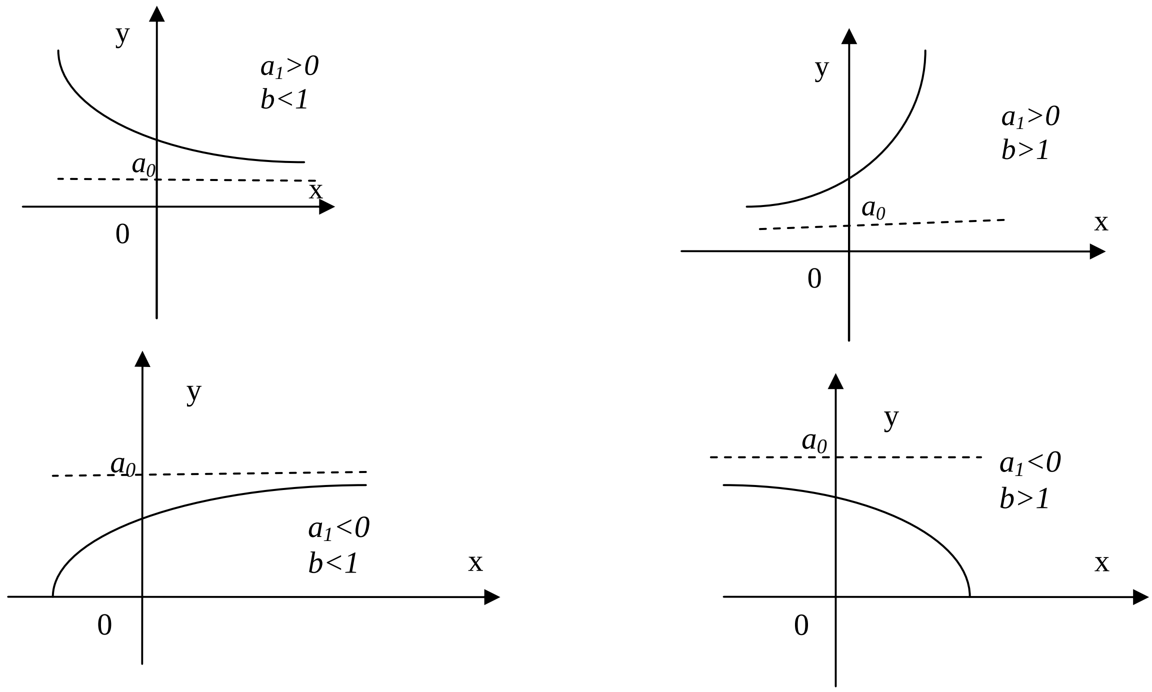


Рис 5.4 – модифицированная показательная зависимость

5. Показательная зависимость с нелинейным показателем.

$$y = b^{a_0 + a\ x + a_2x^2 + ... + a_nx^n} = c_0 \cdot c_1{}^x \cdot c_2{}^{x^2} ... c_n{}^{x^n} \qquad (5.4.6)$$

В частном случае, когда $n{=}2$, имеем

$$y{=}\ c_0\ {\cdot}c_1{}^x{\cdot}c_2{}^{x^2} \qquad (5.4.7)$$

Эта кривая получила название логарифмической параболы. В самом деле, прологарифмировав выражение (5.4.7) , получим:

$$log\ y{=}\ \ log\ (c_0) + x\ log\ (c_1){+}\ x^2\ log\ (c_2) \qquad (5.4.8)$$

6. Логарифмическая зависимость с линейным аргументом

$$y = a{\cdot}\ log_b\ (x) \qquad (5.4.9)$$

7. Модифицированная логарифмическая зависимость

$$y = a_0 + a_1\ log\ _b\ (x) \qquad (5.4.10)$$

где b – известная константа.

8. Логарифмическая зависимость с полиномиальным аргументом

$$y\ {=}b_0{\cdot}\ log_{b1}\ (a_0{+}a_1x{+}\ a_2x^2{+}\dots\ {+}a_nx^n) \qquad (5.4.11)$$

9. Кривая Гомперца

$$y = d_0 \cdot a_1^{b^x} \tag{5.4.12}$$

Эта кривая не симметрична. Если $log(a_1)<0$, то кривая имеет горизонтальную асимптоту и лежит ниже нее, если $log(a_1)>0$, то – выше нее.

На рис. 5.5. приведены четыре варианта кривой Гомперца.

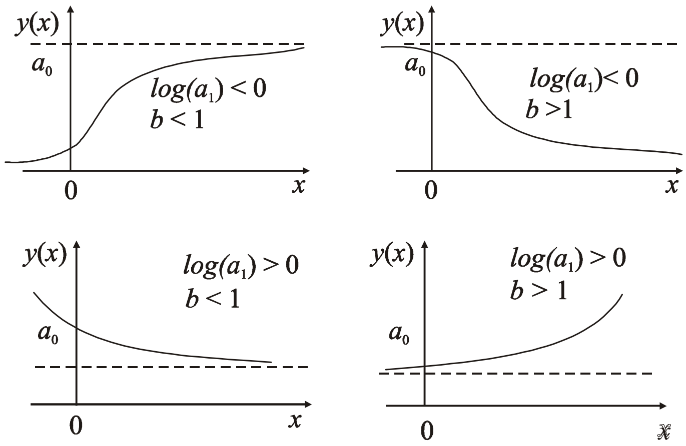


Рис. 5.5

10. Гиперболическая зависимость

$$y = a_0 + \frac{a_1}{x}. \tag{5.4.14}$$

Ясно, что замена $Z = 1/x$ приводит (5.4.14) к виду (5.4.1) $y = a_0 + a_1 z$.

Модель (5.4.14) имеет свои особенности, в отличии от простой линейной регрессии: когда x стремится к бесконечности, величина a_1/x стремится к нулю, а y стремится к предельному значению a_0. Вид модели в значительной мере зависит от знака параметров a_0 и a_1 Рис. 5.6.

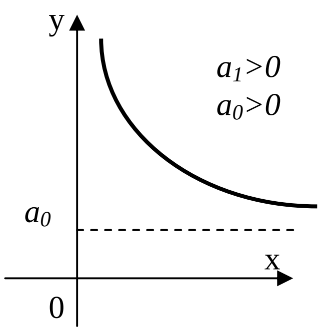


Рис. 5.6.а.

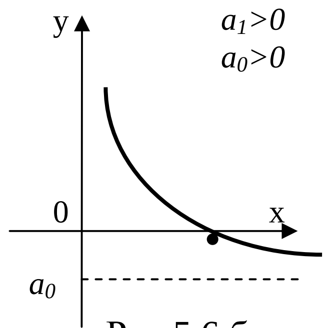


Рис.5.6.б.

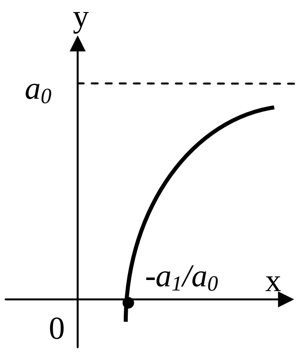


Рис. 5.6.в.

Ярким примером использования гиперболической зависимости в экономиике (в частности в микроэкономике) является известная кривая Филипса (рис. 5.7).

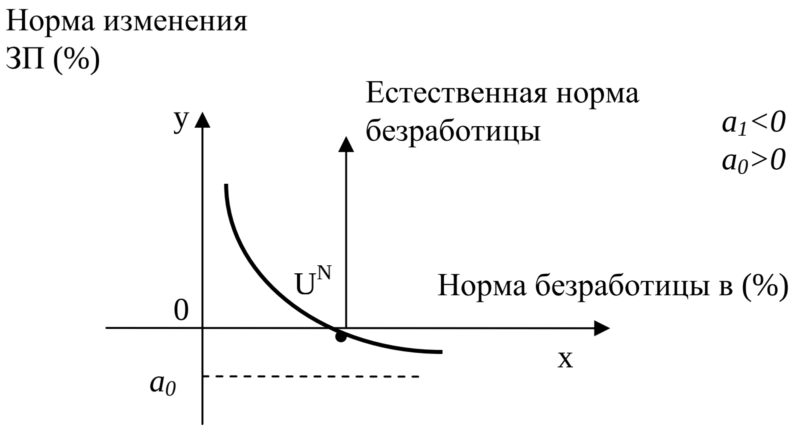


Рис.5.7

Основываясь на данных нормы процента изменения заработной платы и процента безработицы для Англии за период с 1861 по 1975гг., Филипс построил кривую, которая в современной интерпретации изображена на Рис. 5.7. Как показано на рисунке 5.7. асимптота (предельное значение изменения заработной платы) связана с параметром a_0. Точка U^N – значение естественной нормы безработицы; когда $x<U^N$, то норма изменения заработной платы положительная, а когда значение x становится большим, чем естественная норма безработицы, y (норма изменения заработной платы) будет отрицательной. Кривая Филипса дает возможность рассчитать минимальную заработную плату, компенсацию за безработицу и т.д.

Другой важный случай использования гиперболической зависимости – кривая затрат Энгеля, которая связывает потребительские затраты на товары с общими затратами или доходом. Если обозначить через y затраты на потребление, а через x – доход, то кривая Энгеля для определенного товара выявит такие особенности:

а) критический уровень дохода, ниже которого товар не будет куплен равен $-(a_1/a_0)$;

б) граница насыщения, которую нельзя увеличить, как бы ни возрастал доход, равна a_0.

11. Модифицированная гиперболическая зависимость

$$У = \frac{a_0}{1 + a_1 x} \qquad (5.4.15)$$

В этом случае к виду (5.4.1) приводит замена $Z = 1/y$. При этом, $Z = \frac{1}{a_0} + \frac{a_1}{a_0} x = a_0' + a_1' x$

12. Логистическая кривая (обобщение модифицирования гиперболической зависимости).

$y(x) = \frac{a_0}{1 + a_1 \cdot e^{f(x)}}$, где обычно $f(x) = -bx$. Тогда

$$y(x) = \frac{a_0}{1 + a_1 \cdot e^{-bx}} \qquad (5.4.16)$$

Если $b = 1$, а в качестве параметра вместо основания натуральных логарифмов взять основание десятичных логарифмов и принять $f(x) = a_2 x + a_1$, то получим один из наиболее часто используемых видов логистической кривой:

$$y(x) = \frac{a_0}{1 + 10^{a_1 + a_2 x}} \qquad (5.4.17)$$

Типичные графики логистической кривой приведены на рис. 5.8.

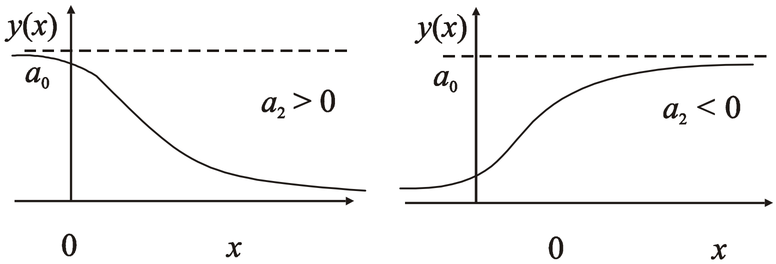


Рис. 5.8

В случае, когда переменная y зависит от нескольких экзогенных $x_1, x_2, \ldots x_n$, для описания этой зависимости используют линейные и нелинейные уравнения регрессии.

13. Линейное многофакторное уравнение регрессии.

$$y = a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n \qquad (5.4.18)$$

14. Нелинейное многофакторное уравнение регрессии.

$$y = a_0 + a_1 x_1 + a_2 x_2 + \ldots + a_n x_n + a_n x^2 + a_{12} x_1 x_2 + a_{13} x_1 x_3 + \ldots + a_{1n} x_1 x_n + a_{22} x^2 + a_{23} x_2 x_3 + \ldots + a_{nn} x^2_n \qquad (5.4.19)$$

Соотношение (5.4.19) является квадратичной формой и может быть записано более компактно: $y = \sum_{i=0}^{n} \sum_{j=0}^{n} a_{ij} x_i x_j$, где $a_{00} = a_0$; $a_{0j} = a_j$; $a_{i0} = a_i$; $x_0 \equiv 1$.

В случае необходимости для описания нелинейных зависимостей могут быть использованы формы более высокого порядка (кубические и т.д.)

5.5. ОЦЕНКА ПАРАМЕТРОВ МОДЕЛЕЙ

Оценивание параметров экономических моделей осуществляется с использованием экспериментальных данных. Как правило, задача оценивания решается следующим образом. Проводится серия экспериментов в каждом из которых фиксируется значения экзогенных переменных $(x_1, x_2,\ldots x_n)$ и соответствующее значения эндогенной переменной y.

Пусть $(x_{i1},x_{i2},\ldots x_{in})$ – набор значений экзогенных переменных в i-ом эксперименте, а y_i – значение эндогенной переменной в этом эксперименте. Тогда, для m экспериментов имеет матрицу значений экзогенных переменных

$X = \begin{pmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{m1} & \cdots & x_{mn} \end{pmatrix}$ и вектор $Y^m= (y_1 y_{2,}\ldots y_m)$ – значение эндогенной переменной.

Теперь необходимо выбрать вид зависимости эндогенной от экзоген-ных. Это чрезвычайно трудная и плохо формализуемая проблема. Как правило, на практике при выборе модели используют соображение здравого смысла, интуицию, аналогию и т.п. Часто осуществляется перебор разных вариантов модели, оцениваются их параметры, а затем окончательно выбирают наиболее адекватную модель. Предположим, что тип модели выбран. Процедура оценивания ее параметров реализуется с использованием какого-либо из статистических методов.

Теорема Гаусса-Остроградского доказывает, что оценки, полученные по методу наименьших квадратов, являются BLUE-оценками, когда выполняются основные предположения относительно случайной величины. Таким образом, оценки, рассчитанные по МНК – линейны, без отклонений, имеют наименьшую дисперсию из всех возможных методов оценки (наилучшая линейная оценка без отклонений).

Сущность МНК рассмотрена на простейшем примере оценивания параметров линейной зависимости. (см. раздел 5.2.1).

Пусть зависимость между эндогенной переменной Y и экзогенной переменной X имеет вид: $y= a_0+a_1x$.

При этом, матрица условий проведения эксперимента X для n экспериментов вырождается в столбец: $x=(x_1, x_2, \dots, x_n)^T$.

Задачу оценивания неизвестных параметров a_0 и a_1 модели (5.4.1) сформулируем следующим образом: найти a_0 и a_1, минимизирующие сумму квадратов отклонений значений эндогенной переменной, предсказываемых в эксперименте. Минимизируемый функционал имеет вид:

$$I=\sum_{j=1}^{n} [y_i-(a_0+a_ix_j)^2] \tag{5.5.1}$$

Значения a_0 и a_1, минимимизирующие (5.5.1), найдем, взяв производные по этим переменным, прировняв их к нулю и решив соответствующие уравнения. В соответствии с этим получим:

$$\begin{cases} \dfrac{\partial J}{\partial a_0} = -2\sum_{j=1}^{n}\left(y_j - a_0 - a_1x_j\right) = 0, \\ \dfrac{\partial J}{\partial a_1} = -2\sum_{j=1}^{n}\left(y_j - a_0 - a_1x_j\right)x_j = 0. \end{cases} \quad \text{или} \quad \begin{cases} a_0 n + a_1\sum_{j=1}^{n} x_j = \sum_{j=1}^{n} y_j, \\ a_0\sum_{j=1}^{n} x_j + a_1\sum_{j=1}^{n} x_j^2 = \sum_{j=1}^{n} x_j y_j. \end{cases} \tag{5.5.2}$$

Решая систему уравнений (5.5.2) методом Крамера относительно a_0 и a_1, получим:

$$a_0 = \frac{\det\begin{pmatrix} \sum_{j=1}^{n} y_j & \sum_{j=1}^{n} x_j \\ \sum_{j=1}^{n} x_j y_j & \sum_{j=1}^{n} x_j^2 \end{pmatrix}}{\det\begin{pmatrix} n & \sum_{j=1}^{n} x_j \\ \sum_{j=1}^{n} x_j & \sum_{j=1}^{n} x_j^2 \end{pmatrix}} = \frac{\left(\sum_{j=1}^{n} x_j^2\right)\left(\sum_{j=1}^{n} y_j\right) - \left(\sum_{j=1}^{n} x_j\right)\left(\sum_{j=1}^{n} x_j y_j\right)}{n\sum_{j=1}^{n} x_j^2 - \left(\sum_{j=1}^{n} x_j\right)^2},$$

$$a_1 = \frac{\det\begin{pmatrix} n & \sum_{j=1}^{n} y_j \\ \sum_{j=1}^{n} x_j & \sum_{j=1}^{n} x_j y_j \end{pmatrix}}{\det\begin{pmatrix} n & \sum_{j=1}^{n} x_j \\ \sum_{j=1}^{n} x_j & \sum_{j=1}^{n} x_j^2 \end{pmatrix}} = \frac{n\left(\sum_{j=1}^{n} x_j y_j\right) - \left(\sum_{j=1}^{n} x_j\right)\left(\sum_{j=1}^{n} y_j\right)}{n\sum_{j=1}^{n} x_j^2 - \left(\sum_{j=1}^{n} x_j\right)^2},$$

Соотношения (5.5.2) заметно упрощаются, если значения эндогенной переменной в эксперименте определяются на системе равноотстоящих значений экзогенной переменной X. В этом случае при надлежащем выборе масштаба можно считать, что $x_1 = \Delta$; $x_2=2\cdot\Delta$; …; $x_n = \Delta\cdot n$. Тогда:

$$\sum_{j=1}^{n} x_j = \Delta\sum_{j=1}^{n} j = \frac{\Delta n(n+1)}{2}; \quad \sum_{j=1}^{n} x_j^2 = \Delta^2\sum_{j=1}^{n} j^2 = \frac{\Delta^2 n(n+1)(2n+1)}{6}.$$

$$a_0 = \frac{\dfrac{\Delta^2 n(n+1)(2n+1)}{6}\left(\sum_{j=1}^{n} y_j\right) - \dfrac{\Delta^2 n(n+1)}{2}\left(\sum_{j=1}^{n} j\cdot y_j\right)}{\dfrac{\Delta^2 n^2(n+1)(2n+1)}{6} - \dfrac{\Delta^2 n^2(n+1)^2}{4}} = \frac{2(2n+1)\sum_{j=1}^{n} y_j - 6\sum_{j=1}^{n} j\cdot y_j}{n(n-1)},$$

$$a_1 = \frac{\Delta n\left(\sum_{j=1}^{n} j\cdot y_j\right) - \dfrac{\Delta n(n+1)}{2}\left(\sum_{j=1}^{n} y_j\right)}{\dfrac{\Delta^2 n^2(n+1)(2n+1)}{6} - \dfrac{\Delta^2 n^2(n+1)^2}{4}} = \frac{12\sum_{j=1}^{n} j\cdot y_j - 6(n+1)\sum_{j=1}^{n} y_j}{\Delta n(n^2-1)}.$$

Например, если используется зависимость (5.4.3), то при логарифмировании правой и левой частей по любому основанию c $(c>0, c\neq 1)$, получим:

$$log_c(y)=log_c(a_0)+xlog_c(b) \qquad (5.5.3)$$

Если теперь обозначить $log_c(a_0)= a_0$, $log_c(b)=b'$, $log_c(y)=y'$, то (5.5.3) примет вид $y= a'_0+a'_1x$, где $a_0 = c^{a'_0}, a_1 = c^{a'_1}$.

При использовании показательной зависимости вида (5.4.6) логарифмирование дает:

$$log_c(y)=log_c(a_0)+xlog_c(c_1)+...+x^k log_c(c_k).$$

После преобразований $log_c(y)=y'$; $log_c(c_0)=c'_0,...,log_c(c_k)=c'_k.$, получим:

$$У'=c'_0+c'_1x+c'_2x^2+\ldots+c'_к x^к$$

и после потенцирования получим искомый набор параметров $(c_1, c_2, \ldots, c_к)$.

В тех случаях, когда линеаризация невозможна, для оценки параметров используется непосредственная минимизация функционала $J = \sum_{j=1}^{m}(y_j - \hat{y}_j)^2$ любым численным методом.

5.6. ПРИМЕРЫ ПОСТРОЕНИЯ ЛИНЕЙНОЙ И ПАРАБОЛИЧЕСКОЙ ЭМПИРИЧЕСКИХ ЗАВИСИМОСТЕЙ

Пример 1. Опытные данные приведены в таблице 1.

Таблица 1.

n	1	2	3	4	5	6	7	8	9	10	11	12	13
x_i	41	50	81	104	120	139	154	180	208	241	250	269	301
y_i	4	8	10	14	16	20	19	23	26	30	31	36	37

По методу наименьших квадратов подобрать и построить прямую, изображающую зависимость y(x).

Решение. Эмпирическая прямая имеет вид: $y=ax+b$.

Для данного частного случая система (5.5.2) будет иметь вид:

$$\begin{cases} \dfrac{\sum_{i=1}^{13} x_i^2}{13}\cdot a+\dfrac{\sum_{i=1}^{13} x_i}{13}\cdot b=\dfrac{\sum_{i=1}^{13} x_i y_i}{13}, \\ \dfrac{\sum_{i=1}^{13} x_i}{13}\cdot a \quad + \quad b=\dfrac{\sum_{i=1}^{13} y_i}{13} \end{cases}$$

Откуда получим:

$$a=\frac{\dfrac{\sum_{i=1}^{13} x_i y_i}{13}-\dfrac{\sum_{i=1}^{13} x_i}{13}\cdot\dfrac{\sum_{i=1}^{13} y_i}{13}}{\dfrac{\sum_{i=1}^{13} x_i^2}{13}-\left(\sum_{i=1}^{13}\dfrac{x_i}{13}\right)^2},\quad b=\frac{\sum_{i=1}^{13} y_i}{13}-a\cdot\frac{\sum_{i=1}^{13} x_i}{13}.$$

Пример 1. (расчёт и график в «EXEL»)

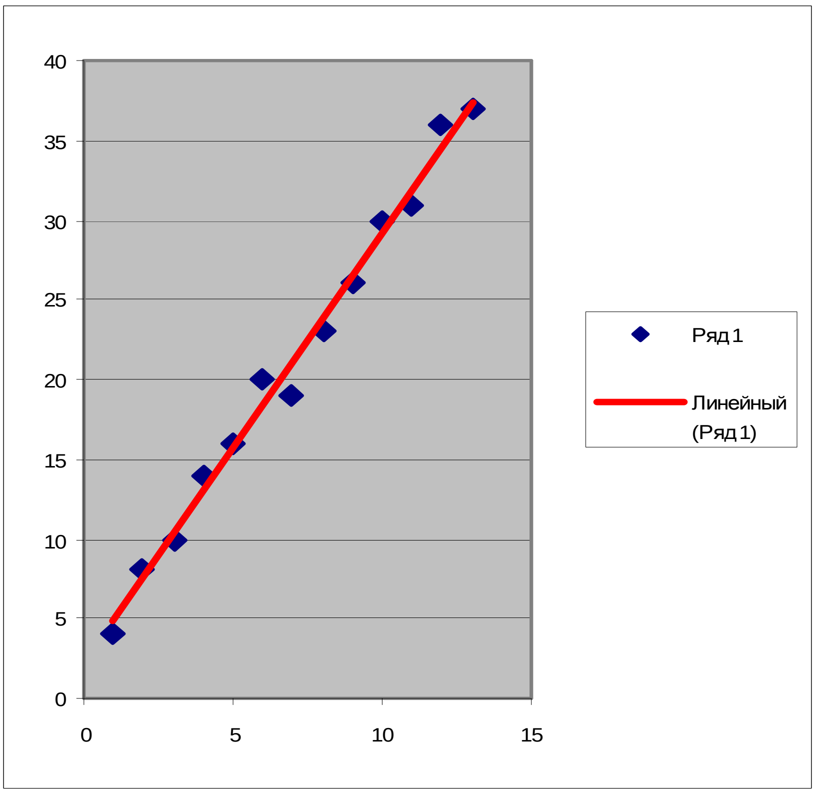


Рис. 5.9

X_i	Y_i	X_i^2	X_iY_i	Y_t
41	4	1681	164	5,758826
50	8	2500	400	6,875472
81	10	6561	810	10,7217
104	14	10816	1456	13,57535
120	16	14400	1920	15,5605
139	20	19321	2780	17,91786
154	19	23716	2926	19,77894
180	23	32400	4140	23,00481
208	26	43264	5408	26,47882
241	30	58081	7230	30,57319
250	31	62500	7750	31,68984
269	36	72361	9684	34,0472
301	37	90601	11137	38,0175
2138	274	438202	55805	
164,4615	21,07692	33707,85	4292,692	
a=0,1241	*b=0,6719*			

Ответ: $y=0,1241x+0,6719$

Пример 2. Опытные данные приведены в таблице 2.

Таблица 2.

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14
xi	1,2	1,31	1,4	1,61	1,74	1,8	2	2,14	2,19	2,41	2,5	2,68	2,81	3
yi	0,54	0,59	0,67	0,76	0,85	0,97	1,07	1,18	1,27	1,39	1,53	1,6	1,78	2,03

По методу наименьших квадратов подобрать и построить параболу, изображающую зависимость *y(x)*.

Решение. Эмпирическая кривая имеет вид:

$y=ax^2+bx+c.$

Пример 2 (расчёт и график в «EXEL»).

Параболическая аппроксимация.				*y=ax^2+bx+c*			
Таблица 2							
Xi	Yi	Xi^2	Xi^3	Xi^4	Xi^2*Yi	Xi*Yi	Yemp
1,2	0,54	1,44	1,728	2,0736	0,7776	0,648	0,357432
1,31	0,59	1,7161	2,248091	2,9449992	1,012499	0,7729	0,485329
1,4	0,67	1,96	2,744	3,8416	1,3132	0,938	0,586288
1,61	0,76	2,5921	4,173281	6,7189824	1,969996	1,2236	0,808962
1,74	0,85	3,0276	5,268024	9,1663618	2,57346	1,479	0,93776
1,8	0,97	3,24	5,832	10,4976	3,1428	1,746	0,994872
2	1,07	4	8	16	4,28	2,14	1,1746
2,14	1,18	4,5796	9,800344	20,972736	5,403928	2,5252	1,290666
2,19	1,27	4,7961	10,50346	23,002575	6,091047	2,7813	1,330173
2,41	1,39	5,8081	13,99752	33,734026	8,073259	3,3499	1,491847
2,5	1,53	6,25	15,625	39,0625	9,5625	3,825	1,552275
2,68	1,6	7,1824	19,24883	51,58687	11,49184	4,288	1,663183
2,81	1,78	7,8961	22,18804	62,348395	14,055058	5,0018	1,735033
3	2,03	9	27	81	18,27	6,09	1,8276
28,79	16,23	63,4881	148,3566	362,95025	88,017187	36,8087	16,23602
2,056429	1,159286	4,534864	10,5969	25,925018	6,286941929	2,629193	1,159716
*25,925*a+10,597*b+4,535*c=6,29*							
*10,597*a+4,535*b+2,056*c= 2,63*				matrix([[-.2047186], [1.6765265], [-1.3595632]])			
*4,53*a+ 2,056*b+ c=1,16*							
a=	-0,2047	*b=*	1,6765	*c=*	-1,3596		

Ответ. $\boxed{y= -0,204x^2+1,6765x-1,3596}$

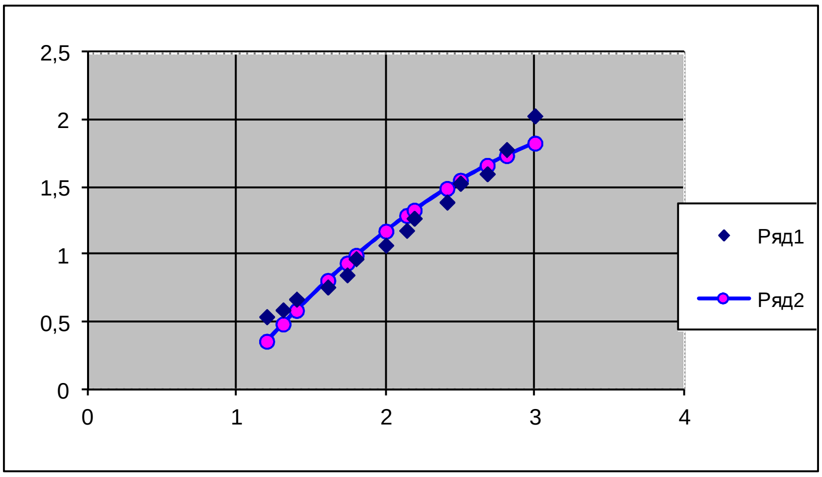


Рис. 5.10

Ряд 1– экспериментальные данные

Ряд 2– парабола: y=-0,204x^2+1,6765x-1,3596

5.7. ПРИМЕРЫ ЗАВИСИМОСТЕЙ, КОТОРЫЕ ЗАМЕНОЙ ПЕРЕМЕННЫХ СВОДЯТСЯ К ЛИНЕЙНЫМ

Легко определяются параметры таких зависимостей, которые следующими подстановками сводятся к линейной зависимости:

1. $y=ax^b$ *(Y=A+BX, где:* $X=\ln x$, $Y=\ln y$, $A=\ln a$, $B=b$*);*
2. $y=ab^{\lambda x}$ *(Y=A+BX, где:* $X=x$, $Y=\ln y$, $A=\ln a$, $B=\lambda \ln b$*);*
3. $y=a+b/x$ *(Y= Ax+B, где:* $X=x$, $Y=xy$, $A=a$, $B=b$*);*
4. $y=1/(ax+b)$ *(Y= Ax+B, где:* $X=x$, $Y=1/y$, $A=a$, $B=b$*);*
5. $y=x/(ax+b)$ *(Y= Ax+B, где:* $X=x$, $Y=x/y$*),* $A=a$, $B=b$*);*
6. $y=a\ln x+b$ *(Y= Ax+B, где:* $X=\ln x$, $Y=y$, $A=a$, $B=b$*).*

Точки, построенные по парам значений X,Y в прямоугольной системе координат, должны расположиться вдоль прямой линии, если имеет место исходная зависимость между переменными x и y. Если же точки «не спрямляются», зависимость выбрана неверно и следует пересмотреть выбор вида эмпирической зависимости. Во всех линейных зависимостях после нахождения A и B, необходимо сделать обратную замену, приводящую к исходным переменным x и y , а значит к выбранному виду связи.

Пример 3. Согласно теоретическим данным зависимость напряжения от времени подчинена закону: $U=U_0 e^{-\lambda t}$. Основываясь на опытных данных (табл. 3), подобрать по методу наименьших квадратов значения параметров U_0 и λ.

Таблица 3.

N_i	1	2	3	4	5	6	7	8	9	10	11
T_i	0	1	2	3	4	5	6	7	8	9	10
U_i	100	75	55	40	30	20	15	10	10	5	5

Решение.

Сделаем замену переменных и сведём к линейной зависимости $Y=AX+B$, где: $X=t$; $Y=\ln U$; $A=\ln U_0$; $B=-\lambda$.

Пересчитаем исходные данные в новых переменных и оформим в виде таблицы 4.

Таблица 4.

N_i	1	2	3	4	5	6	7	8	9	10	11
T_i	0	1	2	3	4	5	6	7	8	9	10
U_i	4,605	4,317	4,007	3,689	3,401	2,996	2,708	2,303	2,303	1,609	1,609

Таким образом, решение этой задачи свелось к линейной аппроксимации, которая рассмотрена в примере 1 в системе EXEL. Просто введем новые данные в готовый алгоритм и получим результат.

Вернёмся к старым переменным и результат изобразим графически.

n	x	y	Y	Xi^2	Xi*Yi	Yemp	Uemp
1	0	100	4,60517	0	0	4,613029	100,789
2	1	75	4,317488	1	4,317488	4,300385	73,72817
3	2	55	4,007333	4	8,014666	3,987741	53,93291
4	3	40	3,688879	9	11,06664	3,675097	39,45248
5	4	30	3,401197	16	13,60479	3,362453	28,85989
6	5	20	2,995732	25	14,97866	3,049809	21,11131
7	6	15	2,70805	36	16,2483	2,737165	15,44314
8	7	10	2,302585	49	16,1181	2,424521	11,29681
9	8	10	2,302585	64	18,42068	2,111877	8,263736
10	9	5	1,609438	81	14,48494	1,799233	6,045007
11	10	5	1,609438	100	16,09438	1,486589	4,421985
sum	55	365	33,5479	385	133,3486		
	5	33,18182	3,049809	35	12,1226	lyambda=	0,312644
	A=	-0,31264		B=	4,613029	U0=	100,789

Ответ. $U = 100,789 \cdot e^{-0,3126t}$ *(Y=-0,3126X+4,61303)*

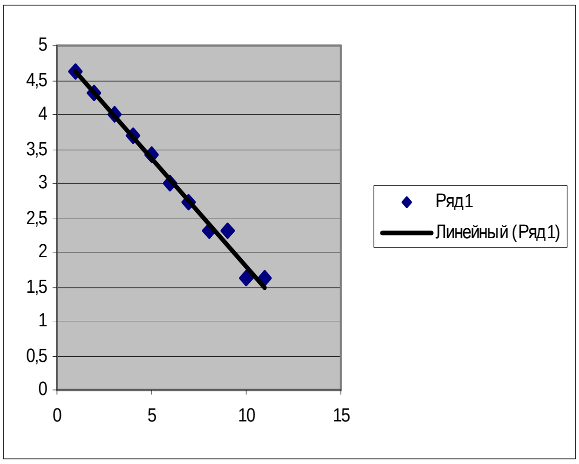


Рис. 5.11.а

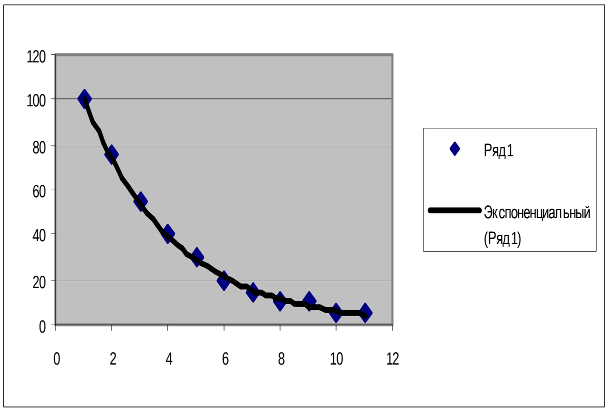


Рис. 5.11.б

РАЗДЕЛ 6. МЕТОДЫ ОПТИМИЗАЦИИ В ЭКОНОМИКЕ

6.1. ОСОБЕННОСТИ И ПРИМЕНЕНИЕ ЭКОНОМИЧЕСКИХ МОДЕЛЕЙ

Современный этап развития страны характеризуется стремлением заменить административные методы управления экономическими.

Один из путей совершенствования методов управления состоит в использовании экономико-математических методов. Необходимость построения математических оптимизационных моделей возникает в связи многочисленностью вариантов создания или функционирования определенной экономической системы, с возможностью применения различного сырья, материалов, технологии для производства одной и той же продукции. Среди этих вариантов по некоторому критерию, отраженному в функции цели, необходимо выбрать наилучший (оптимальный) вариант. Следует также иметь в виду, что множество вариантов функционирования конкретной экономической системы ограничено с точки зрения количества и качества используемого сырья, технологии и т. п.

Процесс построения экономико-математических моделей условно можно разбить на следующие основные этапы.

Первый этап – выбор объекта исследования. В качестве объектов исследования могут быть взяты различные производственно- экономические процессы: размещение производства, перевозка грузов, раскрой промышленных материалов, загрузка производственных мощностей и т.д.

Второй этап – формулировка цели исследования на основе задач, поставленных при изучении данного объекта.

Третий этап – выбор критерия оптимальности, который записывается в виде функционала. Критериями оптимальности обычно служат: максимальный доход, минимальная стоимость исходных материалов, минимальные издержки производства и др.

Четвертый этап – определение основных ограничений. Необходимо при построении моделей выявить основные ограничения и включить их в модель. Как правило, экономические задачи содержат ограничения по ресурсам, поскольку выбор вариантов решения производится в условиях ограниченных ресурсов (оборудование, рабочая сила, сырье, материалы и др.). Кроме ограничений по ресурсам в модель включаются дополнительные условия, определяемые постановкой задачи. К ним относятся, например, соблюдение сроков завершения строительства, удовлетворение спроса на определенную продукцию и др. Система ограничений должна быть достаточно полной, но в то же время не загруженной несущественными ограничениями. Ограничения в модели представляются системами уравнений и неравенств.

Математическое программирование (МП) – раздел математики, относящийся к построению и решению задач на условный экстремум функций многих переменных в заданных областях их изменения.

Математическая модель задачи МП имеет вид: определить экстремум (минимум или максимум) функции $Z=f(x_1,x_2,...,x_n)$ при следующих ограничениях: $g_i(x_1, x_2,..., x_n)\ R\ b_i,\quad i=\overline{1,m}$. Здесь $x_1,x_2,..,\ x_n$ – переменные величины; b_i, $i=\overline{1,m}$ – правые части системы ограничений задачи; выражение $f(x_1,x_2,...,x_n)$ – целевая функция или функция цели, показатель качества решения задачи. В каждом конкретном ограничении вместо * может находиться один из знаков: «=», «≥», «≤».

Целью математического программирования является разработка (там, где это возможно) аналитических методов решения либо создание достаточно эффективных вычислительных процедур отыскания приближенных решений. Целью же решения задачи является выбор программы действий по нахождению оптимальных планов. Отсюда и возникло название отрасли науки — «математическое программирование».

В математическом программировании выделяют два направления. К первому и вполне сложившемуся направлению относят все *детерминированные* задачи, т.е. задачи, у которых вся исходная информация является полностью определенной. Ко второму направлению, называемому *стохастическим* программированием, относятся задачи, в которых либо исходная информация полностью (или частично) является неопределенной, либо некоторые параметры – случайные числа с известными вероятностными распределениями. Так, во многих случаях планирование хозяйственной деятельности происходит без точного знания всей информации о сложившейся ситуации.

По сложившейся традиции в первом направлении математического программирования выделяют следующие классы задач.

1. *Линейное программирование*. Линейное программирование имеет дело с решением задач математического программирования с линейной целевой функцией и линейными ограничениями (равенствами и неравенствами). Это наиболее разработанная отрасль математического программирования. В линейном программировании в свою очередь существуют классы задач, структура которых позволила построить упрощенные методы их решения (например, транспортные задачи).
2. *Целочисленное программирование*. В этом классе задач часть или все переменные должны удовлетворять условию целочисленности.
3. *Параметрическое программирование*. Задачи из данного класса могут содержать коэффициенты, зависящие от параметра.
4. *Нелинейное программирование*. В задачах этого класса нелинейная целевая функция и (или) ограничения. В нелинейном программирова-нии выделяют: 1) выпуклое программирование (нахождение минимума выпуклой функции на выпуклом множестве); 2) квадратичное программирование (нахождение экстремумов квадратичных функций при линейных ограничениях) и 3) многоэкстремальные задачи (например, нахождение минимумов вогнутых функций на выпуклых множествах).
5. *Динамическое программирование*. В этот раздел математического программирования входят методы, основанные на использовании идеи рекуррентного подхода (методы типа математической индукции). Задачи динамического программирования имеют большое число переменных. В них выбор значений переменных производится последовательно, и сделанные выборы относительно просто используются для выбора значений оставшихся переменных. В таких задачах решение одной большой оптимизационной задачи распадается на большое число простых, мелких оптимизационных задач.
6. *Бесконечномерное программирование*. Этот раздел включает в себя методы решения экстремальных задач с бесконечным числом переменных (например, входящие в задачу функции являются функциями, в которых набором переменных является функция (или функции) непрерывного аргумента).

6.2. ЛИНЕЙНАЯ ОПТИМИЗАЦИЯ

Отличительной особенностью задач линейного программирования является то, что функция цели и ограничения являются линейными. *Предмет линейной оптимизации* составляет отыскание экстремума (максимума или минимума) линейной функции при условии, что переменные, подлежащие определению, удовлетворяют линейным ограничениям. Предположение о линейности экономических зависимостей несколько ограничивает возможности линейного программирования, однако простота и наглядность линейных моделей, с достаточной степенью точности описывающих экономические процессы, позволяют применять эти модели в различных видах экономичес-кой деятельности.

Задачи линейного программирования относятся к задачам на условный экстремум функции. Однако, для исследования функции многих переменных на условный экстремум невозможно использование методов математического анализа.

Действительно, пусть необходимо исследовать на экстремум линейную функцию

$Z = c_1x_1 + c_2x_2 + ... + c_nx_n$ при линейных ограничениях:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + ... + a_{1n}x_n = b_1, & x_j \geq 0, \quad j = \overline{1,n}, \\ a_{21}x_1 + a_{22}x_2 + ... + a_{2n}x_n = b_2, & b_i \geq 0, \quad i = \overline{1,m}. \\ \dots \quad \dots \quad \dots \quad \dots \quad \dots & \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m. & \end{cases}$$

Поскольку Z – линейная функция, то $\dfrac{\partial Z}{\partial x_j} = c_j \quad (j = 1, 2, ..., n)$, но все коэффициенты линейной функции не могут равняться нулю, следовательно, внутри области, образованной системой ограничений, экстремальные точки не существуют. Они могут быть лишь на границе области, но исследовать точки границы методами математического анализа невозможно, так как частные производные являются постоянными величинами. Для решения задач линейного программирования созданы специальные методы. Линейное программирование используется в экономике, так как исследование зависимостей между величинами, встречающимися во многих экономических задачах, приводит к линейной функции цели с линейными ограничениями.

6.2.1. Формализация задач линейной оптимизации

Построим некоторые математические модели задач планирования и управления, которые сводятся к задачам линейного программирования.

Задача об оптимальном плане выпуска продукции. Пусть для изготовления n видов продукции $P_1, P_2, ..., P_n$ используют m видов сырья $S_1, S_2, ..., S_m$. Запасы сырья, количество единиц сырья, затрачиваемых на изготовление единицы продукции, а также величина прибыли, получаемая от реализации единицы продукции, приведены в табл. 6.1. Здесь a_{ij} – количество единиц i–го сырья, идущего на изготовление j–ой продукции; c_j– величина прибыли, получаемой при реализации j–й продукции. Пусть x_j– количество единиц j–й продукции, которое необходимо произвести. Требуется составить такой план выпуска продукции, чтобы при ее реализации получить максимальную прибыль.

Функцию цели решаемой задачи – максимальную прибыль при реализации продукции выразим как функцию n переменных x_j, $j=\overline{1,n}$.

Построенная линейная функция Z – *функция цели* совместно с системой ограничений образуют *математическую модель* рассматриваемой задачи.

Таблица 6.1.

Вид сырья	Запас сырья	Количество единиц сырья, идущих на изготовление единицы продукции			
		P_1	P_2	...	P_n
S_1	b_1	a_{11}	a_{12}	...	a_{1n}
S_2	b_2	a_{18}	a_{22}	...	a_{2n}
…	…	...	…	…	…
S_m	b_m	a_{m1}	a_{m2}	…	a_{mn}
Прибыль		c_1	c_2	…	c_n

Математическую модель задачи можно представить в следующем виде.

Найти максимальное значение линейной функции $Z = c_1x_1 + c_2x_2 + ... + c_nx_n$ *при ограничениях*:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + ... + a_{1n}x_n \le b_1, & x_j \ge 0, \quad j=\overline{1,n}, \\ a_{21}x_1 + a_{22}x_2 + ... + a_{2n}x_n \le b_2, & b_i \ge 0, \quad i=\overline{1,m}. \\ \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \le b_m. \end{cases}$$

Задача о раскрое материалов. На производство исходное сырье поступает в виде некоторых целых единиц: рулонов, листов и т. д. Чтобы исходное сырье использовать, необходимо его раскроить на заготовки нужной длины, веса, формы. Рассмотрим одну из таких задач.

Пусть на предприятие поступил некоторый полуфабрикат в виде m различных как по количеству единиц a_i, так и по их размерам партий. В каждой партии содержится полуфабрикат только одинакового размера. Этот полуфабрикат необходимо раскроить на требуемые заготовки таким образом, чтобы максимизировать общее число полных комплектов заготовок. Известно, что в один комплект заготовок s-го вида входит k_s штук ($s=\overline{1,l}$). Пусть единицу полуфабриката i-ой партии можно раскроить на заготовки n_i способами, причем при раскрое ее j-м способом $\left(j=\overline{1,n_i}\right)$ заготовок s-го вида получается a_{ijs} штук. Пусть x_{ij} – число заготовок из партии i, которое запланировано раскроить способом j.

Очевидно, общее число заготовок s, которое можно получить по плану раскроя $X = (x_{i1},...,x_{ij},...,x_{mn})$, равно $\sum\limits_{i=1}^{m}\sum\limits_{j=1}^{n_i} a_{ij}x_{ij}$ Это количество заготовок s можно использовать для составления следующего числа комплектов:

$$\frac{\sum\limits_{i=1}^{m}\sum\limits_{j=1}^{n_i} a_{ij}x_{ij}}{k_s}.$$

Очевидно, что общее число Z полных комплектов заготовок равно наименьшему из полученных выше чисел, т.е.

$$Z=\min_{1\le s\le l}\frac{\sum\limits_{i=1}^{m}\sum\limits_{j=1}^{n_i} a_{ij}x_{ij}}{k_s}.$$

Это число Z следует путем надлежащего выбора плана раскроя X максимизировать.

Ограничения задачи разбиваются на две группы:

1) поступивший на предприятие полуфабрикат должен быть использован полностью: $\sum\limits_{j=1}^{n_i} x_{ij}=a_i$, $\left(i=\overline{1,m}\right)$.

2) количества раскраиваемых единиц полуфабриката по каждому из способов должны быть неотрицательными:

$$x_{ij} \ge 0;\ \sum_{j=1}^{n_i} x_{ij} = a_i\ (i = 1, 2, \ldots, m);\ \left(j = \overline{1, n_i} \right).$$

Получена задача нахождения максимина (максимума от минимума) линейной функции при линейных ограничениях. Эту задачу можно свести к следующей задаче линейного программирования.

Найти максимум функции Z при ограничениях

$$\frac{\sum_{i=1}^{m} \sum_{j=1}^{n_i} a_{ij} x_{ij}}{k_s} \ge y\ (s = 1, 2, \ldots, l);\ \ \sum_{j=1}^{n_i} x_{ij} = a_i\ (i = \overline{1, m});$$

$$y \ge 0,\ x_{ij} \ge 0;\ (i = 1, 2, \ldots, m);\ (j = \overline{1, n_i}).$$

Задача оптимальной загрузки невзаимозаменяемых групп технологического оборудования. Данная задача, как и задача об оптимальной загрузке взаимозаменяемых групп оборудования, принадлежит к классу задач оптимального использования производственных мощностей. В ней предполагается, что выполнение операций по каждому из производственных способов осуществляется на станках нескольких качественно различных групп оборудования (например, на токарных, фрезерных, специальных и т. д.) и в пределах данного варианта технологии оно не может быть передано на станки других групп. В таких задачах многовариантность решений существует за счет взаимозаменяемости различных производственно-технологических способов изготовления деталей. Введем следующие обозначения:

1. i – индекс группы производственного оборудования, число которых равно m;
2. j – индекс вида производимой продукции или осуществляемых операций;
3. n – общее число производимых видов продукции; s – индекс производственно-технологического способа, общее число которых равно n_j;
4. a_{ijs} – коэффициент затрат времени обработки детали j-го вида на оборудовании i-й группы по s-му производственно-технологическому способу в станко-часах на одно изделие;
5. A_i – полезный фонд времени в станко-часах по i-й группе оборудования;

6. c_{js} – себестоимость изготовления одного изделия j-го вида; B_j – минимальное задание по выпуску продукции вида j;

7. x_{js} – планируемый выпуск продукции j-го вида s-м способом.

В принятых обозначениях модель задачи оптимальной загрузки мощностей, обеспечивающая выполнение минимальных производственных заданий B_j $(j=\overline{1,n})$ с минимальной суммарной себестоимостью его выполнения, строится следующим образом.

Минимизировать суммарную себестоимость изготовления всех видов продукции, составляющих производственную программу:

$$Z=\sum_{j=1}^{n}\ \sum_{s=1}^{n_j} c_{js}x_{js}$$

при условиях, что:

1) имеет место баланс между потребным и располагаемым фондом времени работы оборудования $\sum_{j=1}^{n}\sum_{s=1}^{n_j} a_{ijs}x_{js}\le A_i$ $(i=\overline{1,m})$;

2) будут выполнены минимальные производственные задания по выпуску продукции $\sum_{s=1}^{n_j} x_{js}\ge B_j$ $(j=\overline{1,n})$;

3) выполняются условия неотрицательности всех переменных $x_{js}\ge 0$; $(s=\overline{1,n_j};\ j=\overline{1,n})$.

Задача оптимальной загрузки взаимозаменяемого оборудования

Данная задача возникает в случаях, когда требуется найти оптимальный вариант использования фонда времени работы станков, которые могут выполнять одинаковые операции, но с различной производительностью. Производственные способы в ней отличаются специфической особенностью: по каждому способу деталь j-го вида производится лишь на одной i-й группе оборудования с затратами станочного времени a_{ij}.Сами же производственные способы изготовления деталей каждого вида включаются в ограничения. При этом интенсивность применения технологии $(i,\ j)$ характеризует производство деталей j-го вида на i-м оборудовании в штуках x_{ij}, а эффективность её использования выражается показателем прибыли $p_{ij}\,\frac{грн}{шт}$ или затрат $c_{ij}\,\frac{грн}{шт}$.

Требуется составить такой план $X=(x_{ij})$ $(i=\overline{1,m})$; $(j=\overline{1,n})$ использования производственных мощностей, который, максимизирует общую прибыль при производстве и реализации продукции $Z=\sum_{j=1}^{n}\ \sum_{i=1}^{m} p_{ij}x_{ij}$ при ограничениях:

1) имеет место баланс между потребным и располагаемым фондом времени по каждой группе оборудования: $\sum_{j=1}^{n_i} a_{ij}x_{ij} \le A_i, (i=\overline{1,m})$;

2) выполняются производственные задания по выпуску продукции всех видов, т.е. $\sum_{i=1}^{m}\ x_{ij}=B_j \quad (j=\overline{1,n})$;

3) все переменные неотрицательны: $x_{ij}\ge 0 \quad (i=\overline{1,m};\ j=\overline{1,n})$.

Задача о назначениях (проблема выбора). Имеется n механизмов (работников), которые нужно распределить на n работ так, чтобы каждый механизм (работник) выполнял одну и только одну работу и чтобы при заданной производительности каждого механизма на каждой из работ суммарный эффект был максимальным.

Пусть $c_{ij}(i,j=\overline{1,n})$ – производительность i-го механизма на j-й работе; x_{ij} – переменная, равная единице, если *i*-й механизм назначен на *j*-ю работу, и равная нулю в противном случае. Требуется отыскать такой план $X=(x_{ij})$, $(i,j=\overline{1,n})$ распределения механизмов по видам работ при котором суммарная производительность всех механизмов максимальна, т.е. достигает максимума функция $Z=\sum_{i=1}^{n}\ \sum_{j=1}^{n} c_{ij}x_{ij}$ при ограничениях:

1) каждый механизм выполняет одну работу $\sum_{i=1}^{n} x_{ij}=1,\ (i=\overline{1,n})$;

2) каждый вид работы выполняется только одним механизмом $\sum_{i=1}^{n}\ x_{ij}=1$

$$x_{ij}=\begin{cases}1, & \textit{если } i-\textit{механизм назначен на } j-\textit{ю работу};\\ 0, & \textit{в противном случае}\end{cases}$$

Основная задача производственного планирования. Для выпуска комплектной продукции имеется n различных технологий. В одном полном комплекте продукции деталей s-го вида k_s штук $(s=\overline{1,l})$.

Для производства деталей используются m видов ингредиентов (различные виды сырья и другие производственные факторы), имеющихся в ограниченных количествах a_i, $(i=\overline{1,m})$. При этом за один производственный цикл по j-ой технологии $(j=\overline{1,n})$ i-й ингредиент расходуется в количестве a_{ij} единиц и деталей s-го вида получается b_{sj} штук. Если x_j – число циклов производства по j-й технологии, то план $X=(x_1, x_2, ..., x_n)$ производства означает интенсивности применения каждой технологии.

По плану $X=(x_1, x_2, ..., x_n)$ i-й ингредиент расходуется в количестве $\sum_{j=1}^{n} a_{ij} x_j$ единиц и деталей s-го вида получается $\sum_{j=1}^{n_i} b_{sj} x_j$ штук, что позволяет их использовать для составления $\dfrac{\sum_{j=1}^{n_i} b_{sj} x_j}{k_s}$ комплектов и, таким образом, число полных комплектов равно $\min\limits_{1 \le s \le l} \dfrac{\sum_{j=1}^{n} b_{sj} x_j}{k_s}$.

Задача заключается в выборе плана $X=(x_1, x_2, ..., x_n)$, максимизирующего это число полных комплектов продукции, при заданных ограничениях на использование имеющихся запасов ингредиентов $\sum_{j=1}^{n} a_{ij} x_j \le a_i$ $i=\overline{1,m}$ и условии неотрицательности переменных $x_j \ge 0$ $j=\overline{1,n}$.

6.2.2. Векторная и матричная записи задач линейной оптимизации

Задачи линейного программирования описываются линейными соотношениями, которые состоят из функции цели и ограничений, состоящих из уравнений и неравенств. Решение систем линейных неравенств довольно громоздко, поэтому от неравенств переходят к равенствам и решают систему линейных уравнений.

В общем случае математическую модель задачи линейного программирования можно представить в следующем виде:

Найти вектор $X=(x_1,x_2,...,x_n)$, минимизирующий (или максимизирующий) линейную форму (функционал)

$$Z = c_1x_1 + c_2x_2 + ... + c_nx_n = \sum_{j=1}^{n} c_j x_j \tag{6.2.1}$$

при ограничениях:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + ... + a_{1n}x_n * b_1, & x_j \geq 0, \quad j=\overline{1,n}, \\ a_{21}x_1 + a_{22}x_2 + ... + a_{2n}x_n * b_2, & b_i \geq 0, \quad i=\overline{1,m}. \\ \dots \quad \dots \quad \dots \quad \dots \quad \dots & \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n * b_m. & \end{cases} \tag{6.2.2}$$

Задачи линейного программирования, возникающие на практике, описываются линейными соотношениями, которые состоят из функции цели и ограничений видов уравнений, неравенств со знаком «≤» и неравенств со знаком «≥». Знак * в системе ограничений употреблён вместо одного из указанных знаков и знака « = ».

Рассмотрим утверждения, которые позволяют упростить запись условий задачи (6.2.1)− (6.2.2).

Утверждение 1. Задачу максимизации линейной функции всегда можно свести к задаче минимизации некоторой другой линейной функции при тех же ограничениях.

Справедливость этого утверждения вытекает из следующего очевидного соотношения:

$$\max Z = -min(-Z). \tag{6.2.3}$$

Из него следует, что для сведения решения задачи на максимум к решению задачи на минимум достаточно у целевой функции поменять знак на противоположный и найти минимум полученной линейной формы. Очевидно, что оптимальные решения полученных задач будут совпадать, а связь между экстремальными значениями целевых функции осуществляется по формуле (6.2.3). Аналогичным образом в случае необходимости можно свести решение задачи на минимум к решению задачи на максимум.

Утверждение 2. Решение линейной системы неравенств и уравнений всегда можно свести к решению некоторой системы линейных уравнений с неотрицательными переменными, поскольку любое действительное число можно представить в виде разности двух неотрицательных чисел.

Определение. Неотрицательную переменную, которая вводится с коэффициентом +1 или -1 в левую часть неравенства, чтобы преобразовать его в равенство, назовем дополнительной (или свободной) переменной. Дополнительные переменные вводятся в целевую функцию задачи с нулевыми коэффициентами.

Приведенные утверждения позволяют свести решение любой линейной задачи к решению задачи минимизации функции цели, имеющей неотрицательные переменные и ограничения, определяемые системой линейных уравнений. Это даст возможность канонизировать задачи линейного программирования.

Канонической (основной, стандартной, общей) задачей линейного программирования называется задача минимизации линейной формы при условии неотрицательности всех переменных и основных ограничениях-равенствах.

Математическая модель линейной задачи в *канонической форме* записывается следующим образом:

Найти минимум функционала

$$Z = c_1x_1 + c_2x_2 + \ldots + c_nx_n \to min \qquad (6.2.4)$$

при ограничениях

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = b_1, & x_j \ge 0, \quad j = \overline{1,n}, \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = b_2, & b_i \ge 0, \quad i = \overline{1,m}. \\ \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots & \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m. & \end{cases} \qquad (6.2.5)$$

Здесь c_j –коэффициенты целевой функции, a_{ij} ($i = \overline{1,m}$, $j = \overline{1,n}$) — коэффициенты при переменных в основных ограничениях задачи, b_i $(i = \overline{1,m})$ правые части ограничений (6.2.5), а x_j $(j = \overline{1,n})$ – переменные задачи. Эта задача имеет m основных ограничений и n переменных.

Ограничения, связывающие не менее двух переменных, называются основными, а ограничения вида $x_j \ge 0$— условиями неотрицательности переменных. Не ограничивая общности, можно в дальнейшем считать, что правые части $b_i\ (i=\overline{1,m})$ основных ограничений задачи являются неотрицательными числами, в противном случае обе части уравнения, для которого это условие не выполняется, можно умножить на -1.

Задача (6.2.4)– (6.2.5) записана в так называемой развернутой форме. Эта форма удобна при непосредственном решении конкретной задачи.

Используются и другие формы записи канонической задачи:

а) форма записи, использующая знак суммы:

$$Z=\sum_{j=1}^{n} c_j x_j \to \min \qquad \text{при} \qquad \sum_{j=1}^{n} a_{ij} x_j = b_i\,, \quad i=\overline{1,m}; \quad x_j \ge 0,\ j=\overline{1,n};$$

б) матричная форма записи: при $A\vec{X}=\vec{B},\quad \vec{X}\ge 0;$

в) векторная форма записи: $Z=\vec{C}\cdot\vec{X}\to\min,$ при

$$\sum_{j=1}^{n} x_j \overrightarrow{A_j} = \vec{B}, \quad x_j \ge 0,\ j=\overline{1,n}.$$

Здесь $\vec{X}^T=(x_1,x_2,...x_n)$– вектор-столбец переменных величин; $\vec{C}=(c_1,c_2,...c_n)$ – вектор-строка коэффициентов функции цели;

$$A=\begin{pmatrix} a_{11} & a_{12} & ... & a_{1n} \\ a_{21} & a_{22} & ... & a_{2n} \\ ... & ... & ... & ... \\ a_{m1} & a_{m2} & ... & a_{mn} \end{pmatrix}$$ – матрица системы основных ограничений задачи;

$\overrightarrow{A_j}^T=(a_{1j},a_{2j},...a_{nj})$ – *j-й* столбец матрицы *A*; $\vec{B}^T=(b_1,b_2,...b_m)$— вектор-столбец правых частей основных ограничений задачи.

Приведем сначала ряд основных определений теории линейного программирования.

Определение 1. Допустимым решением (планом) задачи линейного программирования называется вектор $X=(x_1,x_2,...,x_n)$, удовлетворяющий всем ее ограничениям (6.2.5).

Определение 2. Допустимое решение (план) $X=(x_1,x_2,...,x_n)$ называется базисным (опорным), если векторы $\vec{A}_j$, входящие в разложение вектора правых частей $\vec{B}$ с положительными коэффициентами x_j являются линейно независимыми. Поскольку векторы $\vec{A}_j$ являются m-мерными, число положительных компонент допустимого базисного решения не может превышать m.

Определение 3. Допустимое базисное решение (опорный план) называется невырожденным, если оно содержит ровно m положительных компонент. В противном случае решение вырожденное.

Определение 4. Оптимальным решением (оптимальным планом) задачи линейного программирования является такое допустимое ее решение (план), которое оптимизирует целевую функцию.

Определение 5. Базисом допустимого базисного решения задачи называется система из m линейно независимых векторов, включающая все векторы $\vec{A}_j$, отвечающие положительным компонентам допустимого базисного решения.

Определение 6. Координаты допустимого базисного решения, отвечающие векторам его базиса, называются базисными, а остальные компоненты – небазисными переменными.

Решение задач линейного программирования основывается на следующих теоремах.

Теорема 1. Множество L всех допустимых решений задачи линейного программирования является выпуклым.

Множество L может быть допустимым либо выпуклым многогранником (ограниченным), либо выпуклой многогранной областью, уходящей на бесконечность (неограниченным).

Теорема 2. Любую точку X ограниченной области L можно представить в виде выпуклой линейной комбинации ее крайних точек.

Очевидно, что любая неограниченная область L допустимых решений задачи линейного программирования имеет конечное число крайних точек, однако не все ее точки можно представить в виде выпуклых комбинаций этих крайних точек. Впредь будем предполагать, что область L, определяемая условиями задачи, ограничена, т.е. совпадает с некоторым многогранником.

Теорема 3. Целевая функция задачи линейного программирования (6.2.4)– (6.2.5) достигает своего экстремума в крайней точке выпуклой области L, являющейся множеством допустимых решений этой задачи.

Теорема 4. *Если линейный функционал достигает экстремума в нескольких крайних точках одновременно, то это же значение он* *принимает в любой точке L, принадлежащей выпуклой оболочке указанных крайних точек.*

Это означает, что если экстремальное значение достигается более чем в одной точке, например, в двух угловых точках, то оно же будет получено в любой точке, являющейся линейной комбинацией данных точек (т.е. на отрезке, соединяющем эти точки).

Сформулируем теорему, которая помогает установить связь между множеством допустимых базисных решений задачи линейного программирования и системой крайних точек соответствующего выпуклого множества L.

Теорема 5. *Функция цели разрешимой линейной задачи достигает своего экстремума хотя бы в одной из угловых точек области L (на одном из допустимых базисных решений системы ограничений).*

Следствие. С каждой крайней точкой L связано ровно m линейно независимых векторов из данной системы n векторов $(A_1, A_2, ..., A_n)$ при условии, что система содержит m линейно независимых векторов.

Эти результаты показывают, что в процессе решения задачи достаточно исследовать лишь те ее допустимые решения, которым соответствует система из m линейно независимых векторов. Так как в данной системе из n векторов содержится не более чем C_n^m различных подсистем, каждая из которых состоит из m линейно независимых векторов, то C_n^m является верхней границей числа допустимых базисных решений задачи линейного программирования, имеющей размеры $m \times n$ (m – основных ограничений, n – переменных). Эти рассуждения позволяют предложить один из возможных путей нахождения оптимального решения – метод сравнения всех базисных решений задачи. Очевидно, уже при сравнительно небольших m и n такой метод является практически неприемлемым.

В связи с этим возникает необходимость в таком методе решения задачи, который упорядочивал бы переход от одного допустимого базисного решения к другому и позволял бы получить оптимальное решение за сравнительно небольшое конечное число шагов. Таким методом является симплекс-метод, разработанный Дж. Данцигом, а также различные его модификации. Этот метод гарантирует получение оптимального решения не более чем за $2m$ шагов. За это же число шагов метод устанавливает отсутствие допустимых решений задачи либо неограниченность целевой функции на множестве допустимых решений.

6.2.3. Графический метод решения задач линейной оптимизации

Рассмотрим задачу нахождения экстремума линейной функции

$$Z = \sum_{j=1}^{n} c_j x_j \tag{6.2.6}$$

при условиях

$$\sum_{j=1}^{n} a_{ij} x_j = b_i\,,\ i = \overline{1,m}\,, \tag{6.2.7}$$

$$x_j \geq 0,\ j = \overline{1,n}; \tag{6.2.8}$$

Каждое из уравнений системы (6.2.7) определяет в n-мерном пространстве некоторую гиперплоскость, представляющую собой выпуклое множество. Каждое неравенство (6.2.8) определяет в том же пространстве выпуклое множество (полупространство). Совокупность всех допустимых решений L системы ограничений (6.2.7) – (6.2.8) является пересечением m гиперплоскостей и n полупростанств, т. е. является выпуклым множеством.

Следовательно, решение задачи (6.2.7) – (6.2.8) состоит в определении экстремума линейной функции на некотором выпуклом множестве. Если это множество ограничено, то оно представляет собой многогранник в n-мерном пространстве. В общем случае совокупность допустимых решений задачи (6.2.7) – (6.2.8) может быть плоскостью, полупространством, неограниченным и ограниченным многогранником, многоугольником, прямой, полупрямой, отрезком, точкой, пустым множеством.

Выясним геометрический смысл целевой функции (6.2.6). Для этого используем понятие линии (поверхности) уровня функции. Напомним, что линия (поверхность) уровня функции – это совокупность всех точек пространства, в каждой из которых данная функция принимает постоянное значение.

Рассмотрим линейное уравнение

$$\sum_{j=1}^{n} c_j x_j = b \tag{6.2.9}$$

при произвольном b.

Для линейных функций все линии (поверхности) уровня параллельны, так как представляют собой либо прямые на плоскости, либо плоскости, либо гиперплоскости ($n \geq 4$). Такие гиперплоскости (6.2.9) используются для решения вопроса об отыскании в области *L* экстремальных точек целевой функции и поэтому часто называются разрешающими. Очевидно, экстремальными точками области *L* (если экстремумы целевой функции достигаются) будут те ее точки, которые отвечают так называемым предельным положениям разрешающей гиперплоскости. Назовем предельным положением разрешающей гиперплоскости такое ее положение, при котором она проходит хотя бы через одну точку области *L*, но не содержит ни одной ее внутренней точки.

Тот факт, что экстремальные точки целевой функции соответствуют предельным положениям разрешающей гиперплоскости, подтверждается свойством градиента линейной целевой функции (6.2.6), который указывает направление наискорейшего возрастания этой функции (антиградиент функции в данной точке указывает направление наискорейшего убывания функции). Градиент функции $Z = f(x_1, x_2, ..., x_n)$ в точке $\overline{X} = (\overline{x}_1, \overline{x}_2 ..., \overline{x}_n)$ – это вектор, координаты которого равны значениям соответствующих частных производных этой функции в данной точке, т. е.

$$grad\ f(\overline{X}) = \nabla f(\overline{X}) = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, ..., \frac{\partial f}{\partial x_n} \right)_{\overline{X}}.$$

Градиент функции цели равен $grad\ Z = (c_1, c_2, ..., c_n) = C$.

Таким образом, в каждой точке пространства градиент линейной функции постоянный и совпадает с вектором, составленным из коэффициентов данной функции при переменных. Кроме того, градиент линейной функции перпендикулярен каждой линии (поверхности) ее уровня. Существует взаимно однозначное соответствие между множествами экстремальных точек целевой функции и предельными положениями соответствующей разрешающей гиперплоскости.

Для нахождения предельных положений разрешающей гиперплоскости (плоскости, прямой) нужно построить какую-нибудь линию (поверхность) уровня целевой функции (например, *Z=0)* и, передвигая ее параллельно самой себе в направлении области L, определить экстремальные точки целевой функции.

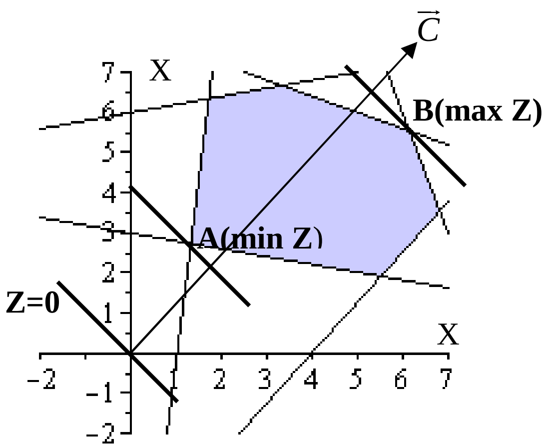


Рис.6.1

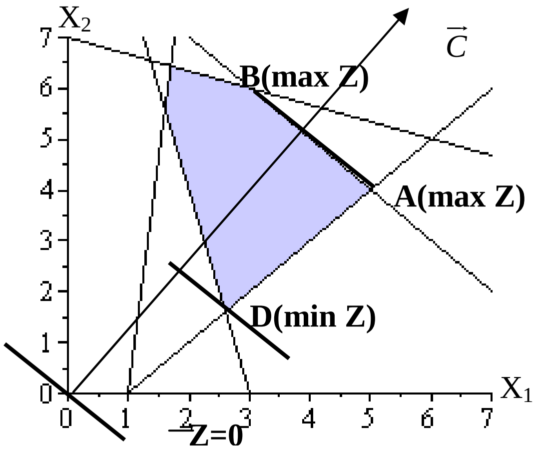


Рис.6.2

Предельное положение разрешающей гиперплоскости, полученное с помощью ее параллельного перемещения как можно дальше (ближе) в направлении градиента (в направлении, противоположном градиенту) целевой функции, отвечает ее максимальному (минимальному) значению (рис.6.1).

Приведенные рассуждения являются основой графического метода решения задачи линейного программирования. Рассмотрим различные возможные случаи взаимного положения разрешающей плоскости (т.е. предельных положений) и многогранника. Разрешающая гиперплоскость в одном из своих предельных положений может оказаться параллельной ребру или грани (рис.6.2). Тогда экстремум линейного функционала достигается в каждой точке такого ребра или грани.

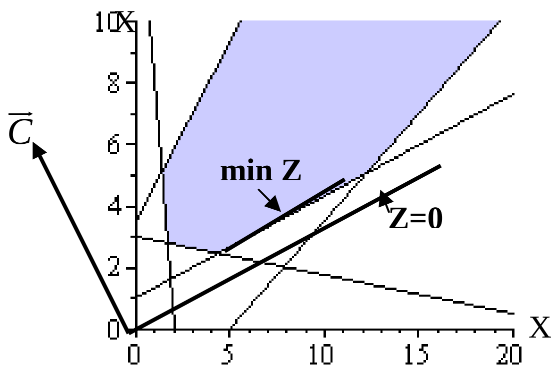


Рис.6.3.

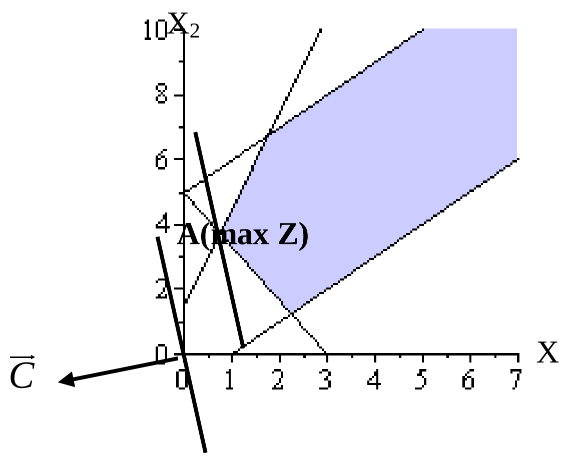


Рис.6.4

Если же множество допустимых решений задачи не ограничено, то целевая функция может оказаться неограниченной как сверху (рис.6.3), так и снизу (рис.6.4). В то же время другой экстремум функции в обоих случаях достигается.

Наличие условий неотрицательности в ограничениях задачи $x_j \geq 0$, $j = \overline{1,n}$ всегда гарантирует ограниченность одного из экстремумов линейного функционала.

Перейдем к непосредственному изложению графического метода решения задач линейного программирования.

Основные идеи этого метода сформулированы выше: построить область допустимых решений L и разрешающую гиперплоскость $Z = 0$; передвигая последнюю параллельно самой себе, определить предельные ее положения. В тех точках L, на которых достигаются эти предельные положения, и будет оптимум целевой функции задачи.

Графический метод целесообразно использовать в: 1) задачах, содержащих не более двух переменных; 2) задачах с числом переменных большим двух, но основные ограничения которых после приведения к стандартному виду содержат не более двух свободных переменных.

Задачи второго типа предварительно должны быть приведены к первому типу. Такое приведение осуществляется с помощью следующих действий:

1) переменные задачи, ограничения которой предварительно приведены к стандартному виду, разделяют с помощью метода полного исключения на свободные и базисные. Для этого полученная система уравнений- ограничений решается методом полного исключения. В результате в случае непротиворечивой системы уравнений получают ее решение (общее или единственное). Если же система уравнений противоречива, то процесс решения задачи заканчивается и задача не имеет решения, т.е. L – пустая область). В противном случае переходят к следующему действию;

2) одновременно исключают базисные переменные из целевой функции;

3) в полученной системе уравнений опускают базисные переменные, т.е. заменяют ее эквивалентной системой линейных неравенств, содержащей только свободные переменные. Эта система неравенств с учетом неотрицательности свободных переменных описывает область допустимых значений этих переменных. Объединяя преобразованную целевую функцию и ограничения на значения свободных пере-менных, получим задачу первого типа, решаемую графическим методом;

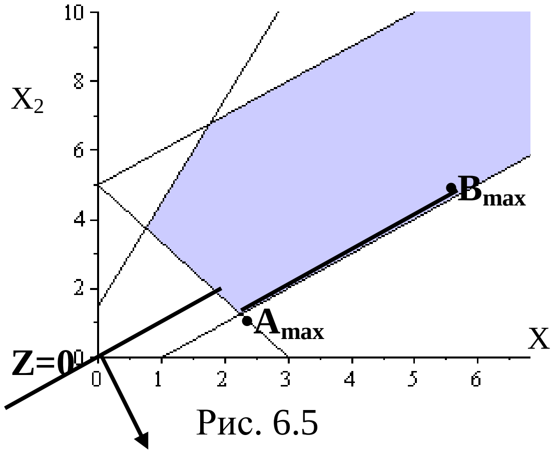


Рис. 6.5

4) задача из предыдущего пункта решается графическим методом. Если целевая функция окажется неограниченной – вычисления заканчиваются, в противном случае переходим к следующему действию;

5) по найденным в пункте 4) оптимальным значениям свободных перемен-ных с помощью общего решения ограничений исходной задачи един-ственным образом определяются значения базисных переменных. Объединяя найденные значения всех переменных, получаем оптимальное решение исходной задачи и соответствующее ему экстремальное значение функции цели.

Если задача разрешима, то она имеет либо одно оптимальное решение, либо бесчисленное их множество (альтернативный оптимум). В случае альтернативного оптимума (рис.6.2) в большинстве случаев есть, по крайней мере, две вершины области L, в которых целевая функция достигает своего экстремума. В этом случае любая точка отрезка AB является экстремальной.

В случае неограниченности области L может оказаться, что среди бесчисленного множества оптимальных решений (рис.6.5) только одно $X_{îïò}$ совпадает с вершиной области.

Графический метод основывается на геометрической интерпретации задачи линейного программирования и возможности наглядного (графического) представления области Q или области допустимых значений свободных переменных этой же задачи. Для построения области допустимых значений Q строится множество решений каждого ограничения задачи, а затем находится их пересечение. Отметим, что линейное неравенство на плоскости определяет полуплоскость, а в пространстве − полупространство; линейное уравнение на плоскости определяет прямую, а в пространстве − плоскость.

Пример.1. Задача об оптимальном плане выпуска продукции

Пусть для изготовления 2-х видов продукции P_1, P_2 используют 2 вида сырья S_1, S_2. Запасы сырья, количество единиц сырья, затрачиваемых на изготовление единицы продукции, а также величина прибыли, получаемая от реализации единицы продукции, приведены в условии задачи. Пусть x_j – количество единиц *j*–й продукции, которое необходимо произвести. Требуется составить такой план выпуска продукции, чтобы при ее реализации получить максимальную прибыль.

Функцию цели решаемой задачи – максимальную прибыль при реализации продукции выразим как функцию 2-х переменных x_j , $j = \overline{1,2}$.

Найдём максимум функции $Z = 6x_1 + 2x_2$ при ограничениях

$$\begin{cases} 5x_1 + 6x_2 \le 30, \\ 7x_1 + 6x_2 \le 42, \\ 0 \le x_2 \le 4, \\ 0 \le x_1 \le 5; \end{cases}$$

Строим область *L* допустимых решений задачи. Для этого сначала построим области решений каждого ограничения − неравенства. Пересечение данных областей определит область *L*.

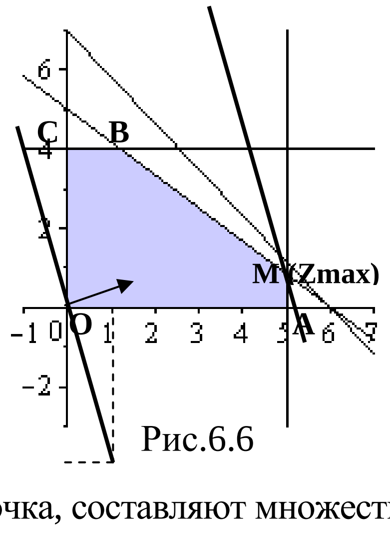


Рис.6.6

В каждом неравенстве задачи заменим знак неравенства на знак равенства. Получим уравнения прямых, ограничивающих соответствующие полуплоскости решений. Для того чтобы определить требуемую полуплоскость, берем любую точку плоскости (например, *O (0, 0)*) и подставляем ее координаты в неравенство. Если неравенство удовлетворяется, то все точки плоскости, лежащие по ту сторону от прямой, что и данная точка, составляют множество решений соответствующего неравенства. Если же неравенство не выполняет-ся, то те точки плоскости, которые лежат в противоположной стороне по от-ношению к выбранной точке и разграничивающей прямой, образуют множество решений данного неравенства.

Область L приведена на рис 6.6. Требуемые полуплоскости отмечаем заливкой. Условия $x_1 \geq 0; x_2 \geq 0$ ограничивают многоугольник решений $OAMBC$ первой четвертью.

Для определения экстремума строим прямую, приравнивая функцию цели некоторому числу (например, нулю) $Z = 6x_1 + 2x_2$=0. Строим вектор градиент $grad\ Z = \vec{C} = (6,2)$ целевой функции и линию уровня функции $Z : 6x_1 + 2x_2 = h;$ для $h = 0$ линия уровня проходит через точку *(0;0)* перпендикулярно вектору $grad\ Z$.Перемещая прямую в направлении вектора $grad\ Z$, получаем, что максимальное значение функционал принимает в вершине M, наиболее удаленной в направлении вектора $gradZ$ вершине многоугольника L. Координаты точки M найдем как точки пересечения прямых (1) и (4): $5x_1 + 6x_2 = 30$, $x_1 = 5$ получим $x_1 = 5; x_2 = 5/6;$ т.е. точка $B(5;5/6)$; $Z_{\max} = 6 \cdot 5 + 2 \cdot 5/6 = 31\,2/3.$

Для ограниченных областей графический метод можно упростить. Можно найти координаты всех угловых точек области L и вычислить в них значение функционала. Наибольшее из этих значений будет максимальным значением Z, а наименьшее $-$ минимальным.

Пример 2. Определить графическим методом минимум функционала.
$Z = -6x_2 - 4x_3 + 2x_4 - x_5$ при ограничениях

$$\begin{cases} 3x_1 - x_2 - 2x_3 + x_4 = 1, \\ -5x_1 + 3x_2 + x_4 + x_6 = 2, \\ -x_1 + x_2 + x_3 = 1, \\ -5x_1 + x_5 + x_6 = 1, \\ x_j \geq 0, \quad j = \overline{1,6} \end{cases}$$

Основные ограничения в задаче являются равенствами, а переменные неотрицательны. Целевую функцию записываем в виде уравнения $Z + 6x_2 + 4x_3 - 2x_4 + x_5 = 0$, присоединяем его к уравнениям системы ограничений и разделяем переменные на свободные и базисные с помощью *метода* *полного исключения*. Базисные переменные одновременно с этим будут исключены из целевой функции.

Z	x_1	x_2	x_3	x_4	x_5	x_6	b
0	3	-1	-2	1	0	0	1
0	-5	3	0	1	0	1	2
0	-1	1	1	0	0	0	1
0	-5	0	0	0	1	1	1
1	0	6	4	-2	1	0	0
0	3	-1	-2	1	0	0	1
0	-5	3	0	1	0	1	2
0	-1	1	1	0	0	0	1
0	-5	0	0	0	1	1	1
1	5	6	4	-2	0	-1	-1
0	3	-1	-2	1	0	0	1
0	-8	4	2	0	0	1	1
0	-1	1	1	0	0	0	1
0	-5	0	0	0	1	1	1
1	1	4	0	0	0	-1	1
0	3	-1	-2	1	0	0	1
0	-8	4	2	0	0	1	1
0	-1	1	1	0	0	0	1
0	3	-4	-2	0	1	0	0
1	3	8	2	0	0	0	2
0	1	1	0	1	0	0	3
0	-6	2	0	0	0	1	-1
0	-1	1	1	0	0	0	1
0	1	-2	0	0	1	0	2
1	5	6	0	0	0	0	0

Приведём эквивалентную задачу:

$$Z=-5x_1-6x_2 \to \min$$

$$\begin{cases} x_1+x_2 \qquad +x_4 = 3, \\ -6x_1+ \quad 2x_2+x_6 = -1, \\ -x_1 \quad +x_2 \quad +x_3 = 1, \\ x_1 \quad -2x_2 \quad +x_5 = 2, \\ x_j \ge 0, \quad j=\overline{1,6} \end{cases}$$

Общее решение системы ограничений задачи: $\begin{cases} x_1 \ge 0, \\ x_2 \ge 0 \\ x_3 = x_1-x_2+1 \ge 0, \\ x_4 = -x_1-x_2+3 \ge 0, \\ x_5 = -x_1+2x_2+2 \ge 0, \\ x_6 = 6x_1-2x_2-1 \ge 0. \end{cases}$

Принимая во внимание условие неотрицательности переменных, получим эквивалентную задачу: найти минимум функционала $Z=-5x_1-6x_2$ при ограничениях:

$$\begin{cases} x_1-x_2 \ge -1, & (1) \\ x_1+x_2 \le 3, & (2) \\ x_1-2x_2 \ge -2, & (3) \\ -6x_1+2x_2 \ge -1, & (4) \\ x_1 \ge 0,\ x_2 \ge 0, & (5) \end{cases}$$

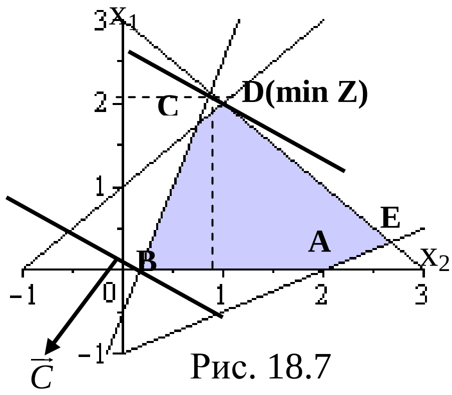


Рис. 18.7

Полученная задача относительно свободных переменных x_1, x_2 решается графическим методом. Многоугольник решений: $ABCDE$. Строим вектор $\vec{C}$=(-5; -6). Передвигая прямую $-5x_1-6x_2=0$ параллельно себе в направлении, противоположном вектору $\vec{C}$, определяем наиболее удаленную в этом направлении точку $D(1;2)$, лежащую на пересечении прямых (1) и (2). $\overrightarrow{X}_{\min}$ = = (1;2), $Z_{\min}$ = *-17*. Значения базисных переменных находим из общего решения: x_3= *1-2+1*, x_4 =*-1-2+3=0*, x_5=*-1+4+2=5*, x_6=*6-4-1=1*.

Следовательно, $\overrightarrow{X}_{\min}$ = *(1; 2; 0; 0; 5; 1)*, $Z_{\min}$ = *−17*.

6.3. СИМПЛЕКСНЫЙ МЕТОД

Как следует из теорем линейного программирования, оптимальное решение задачи линейного программирования достигается в одной из угловых точек многоугольника (многогранника) решений, координаты которой соответствуют базисному решению системы ограничений.

Основная идея симплексного метода. Требуется оптимизировать функционал

$$Z = \sum_{j=1}^{n} c_j x_j \tag{6.3.1}$$

при условиях

$$\sum_{j=1}^{n} a_{ij} x_j = b_i\,, \quad i = \overline{1,m}; \quad x_j \geq 0,\ j = \overline{1,n}; \tag{6.3.2}$$

Ограничения (6.3.2) в n-мерном пространстве определяют некоторый выпуклый многогранник, в одной из крайних точек которого достигается экстремум линейного функционала. Последнее условие позволяет ограничить поиск оптимального решения вершинами многогранника L.

Симплекс-метод включает в себя 1) определение одной из вершин многогранника L (начального опорного плана задачи); 2) направленный перебор вершин L (допустимых базисных решений задачи), при котором на каждом шаге осуществляется приближение к оптимальному решению.

Геометрически это означает, что, если в выбранной вершине не достигнуто экстремальное значение функции цели, осуществляется переход к соседней угловой точке путем изменения базисного решения (замены базиса). Переход происходит упорядоченно (например, в направлении уменьшения функция цели); кроме того, для перехода к соседней вершине достаточно в предыдущем базисном решении заменить лишь одно базисное неизвестное на другое, т.е. происходит однократное замещение базиса.

Таким образом, решение задачи линейного программирования симплекс-методом распадается на ряд шагов. После конечного числа таких шагов мы приходим либо к опорному плану, на котором достигается экстремум функции цели, либо выявляем неограниченность целевой функции, либо устанавливаем неразрешимость задачи.

Имеет место **теорема о конечности алгоритма симплекс - метода.**

Если существует оптимальное решение задачи линейного программирования, то оно может быть получено симплекс-методом за конечное число шагов.

6.3.1. Определение начального план.

Пусть каждое ограничение системы ограничений-равенств имеет переменную, входящую в левую часть с коэффициентом, равным единице, а во все остальные – равным нулю (при неотрицательности правых частей). Говорят, что указанные ограничения имеют предпочтительный вид. Для получения системы ограничений предпочтительного вида достаточно применить метод Жордана-Гаусса. В этом случае можно найти ее опорное решение (базисное с неотрицательными координатами). Все переменные, кроме базисных, нужно приравнять к нулю. Тогда последние примут значения, равные правым частям. Поскольку полученный план будет иметь не более m отличных от нуля координат, то по теореме о структуре координат базисного решения полученный план будет опорным. Переменные, значения которых приравниваем к нулю, будут свободными.

При приведении систем ограничений – неравенств к предпочтительному виду рассмотрим следующие случаи.

1. Пусть система такова:

$$\sum_{j=1}^{n} a_{ij} x_j \le b_i\,, \quad i=\overline{1,m}; \quad b_i \ge 0.$$

Добавим к левым частям *дополнительные переменные* $x_{n+i} \ge 0,\ \ i=\overline{1,m}.$ Получим *расширенную* задачу, эквивалентную исходной. Система ограничений в расширенной задаче имеет предпочтительный вид:

$$\sum_{j=1}^{n} a_{ij} x_j + x_{n+i} = b_i\,, \quad i=\overline{1,m}; \quad b_i \ge 0.$$

Следовательно, начальный опорный план преобразуется:

$$\vec{X}_0 = (\underbrace{0,0,...,0}_{n},\underbrace{b_1,b_2,...,b_m}_{m}).$$

Дополнительные переменные в целевую функцию вводятся с коэффициентами, равными нулю: $c_{n+i} = 0, i = \overline{1,m}.$

2. Рассмотрим систему ограничений

$$\sum_{j=1}^{n} a_{ij} x_j \ge b_i, \quad i = \overline{1,m}; \quad b_i \ge 0.$$

Вычтя из левых частей дополнительные переменные $x_{n+i} \ge 0, \ i = \overline{1,m}$, получим расширенную задачу, эквивалентную исходной. Однако теперь в ней система ограничений не имеет предпочтительного вида, так как дополнительные переменные входят в левую часть с коэффициентами, равными минус единице. Поэтому план

$$\vec{X}_0 = (\underbrace{0,0,...,0}_{n}, \underbrace{-b_1, -b_2, ..., -b_m}_{m})$$

– недопустим. В этом случае вводят так называемый искусственный базис. К левым частям системы ограничений – равенств, не имеющих предпочтитель-ного вида, добавляют *искусственные переменные* ω_i. В целевую функцию ω_i вводят с коэффициентом, равным $(+M)$ в случае решения задачи на *min*, и с коэффициентом, равным *M* в случае решения задачи на *max*, где *M* большое положительное число. Для исходной задачи полученная задача называется M – задачей. Очевидно, она всегда имеет предпочтительный вид.

Рассмотрим исходную задачу линейного программирования:

$$\min(max)\ Z = \sum_{j=1}^{n} c_j x_j \tag{6.3.3}$$

при условиях

$$\sum_{j=1}^{n} a_{ij} x_j = b_i, \quad b_i \ge 0, i = \overline{1,m}; \tag{6.3.4}$$

$$x_j \ge 0. \tag{6.3.5}$$

Если ни одно из ограничений не имеет предпочтительного вида, то M – задача запишется:

$$\min(max)\,\overline{Z}=\sum_{j=1}^{n}c_j x_j-(+)\sum_{i=1}^{m}M\omega_i; \tag{6.3.6}$$

$$\sum_{j=1}^{n}a_{ij}x_j+\omega_i=b_i\,,\quad b_i\ge 0,\, i=\overline{1,m}; \tag{6.3.7}$$

$$x_j\ge 0,\; j=\overline{1,n},\quad \omega_i\ge 0,\quad i=\overline{1,n}, \tag{6.3.8}$$

где знак (+) в (6.3.6) относится к задаче на min. M – задача предпочтительного вида, ее начальный опорный план

$$\overline{\vec{X}}_0=(\underbrace{0,0,...,0}_{n},\underbrace{b_1,b_2,...,b_m}_{m}). \tag{6.3.9}$$

Между оптимальным планом исходной задачи (6.3.3)– (6.3.5) и оптимальным планом M – задачи (6.3.6)– (6.3.8) существует связь. Если в оптимальном плане M – задачи

$$\overline{\vec{X}}^*=(x_1^*,x_2^*,...,x_n^*,\omega_1^*,\omega_2^*,...,\omega_m^*)$$

– все искусственные переменные равны нулю, то план

$$\vec{X}^*=(x_1^*,x_2^*,...,x_n^*)$$

является оптимальным для исходной задачи. Если же хотя бы одна из них не равна нулю, то система ограничений исходной задачи несовместна.

6.3.2. Свойства алгоритма симплекс-метода

Основа алгоритма симплексного метода – получение нового допустимого базисного решения, исходя из известного исходного (выбор вектора, вводящегося в базис известного допустимого базисного решения и определение вектора, выводящегося из базиса).

Теорема. *Если вектор (переменную), вводящийся из базиса (базисного решения) имеющегося допустимого базисного решения, выбирать по наименьшему симплексному отношению, то полученному базису в результате одного шага метода полного исключения можно поставить в соответствие новое допустимое базисное решение.*

Критерий разрешимости линейного программирования

Рассмотрим алгоритм отыскания оптимального решения. Возникает в связи с этим необходимость уметь целенаправленно перебирать допустимые базисные решения, т.е. желательно уметь в качестве последующего выбирать такое допустимое базисное решение, для которого каждый раз значение целевой функции оказывалось бы ближе к оптимальному решению. Предположим, что каждое ее допустимое базисное решение является невырожденным (допустимое решение задачи вырождено, если некоторые из его базисных компонент равны нулю).

Пусть задача линейного программирования разрешима, т.е. множество ее допустимых решений не является пустым. Пусть также известно ее некоторое допустимое базисное решение, $\vec{X}_0 = (x_{l_1}, x_{l_2}, ..., x_{l_m})$, которому соответствуют следующее разложение вектора правых частей и значение целевой функции

$$\sum_{i=1}^{m} x_{li} A_{li} = A_0, \tag{6.3.10}$$

$$\sum_{i=1}^{m} a_{li} x_{li} = Z, \tag{6.3.11}$$

где все $x_{li} > 0$. Каждый вектор A_j первоначальной системы $A_1, A_2, ..., A_n$ разложим по векторам базиса $A_{l1}, A_{l2}, ..., A_{lm}$

$$A_j = \sum_{i=1}^{m} x_{ij} A_{li},\ j = \overline{1,n} \tag{6.3.12}$$

В силу единственности, данного разложения ему можно поставить в соответствие единственное значение линейного функционала Z:

$$\sum_{i=1}^{m} x_{lj} c_{lj} = Z_j,\ j = \overline{1,n} \tag{6.3.13}$$

где c_{li} – коэффициент целевой функции, соответствующий переменной x_{li}.

Образуем так называемые разности $Z_j - c_j$. Условие направленного перехода от одного допустимого базисного решения к другому состоит в выборе такого вектора для введения в базис, которому соответствует положительная разность $Z_j - c_j$. Это утверждение следует из следующей теоремы, которую называют теоремой о возможности построения нового допустимого решения с меньшим значением целевой функции.

Теорема. Если для известного допустимого базисного решения $\vec{X}_0$ при некотором фиксированном j разность $Z_j - c_j > 0$, то можно построить такое множество допустимых решений задачи, что для любого из них $Z < Z_0$, где $Z -$ значение функционала, соответствующее этому допустимому решению.

Предположим, что целевая функция Z ограничена на множестве L. Докажем, что можно построить новое допустимое базисное решение задачи, которому отвечает меньшее значение целевой функции, чем предыдущему (отыскание минимума).

Доказательство. Равенства (6.3.12)– (6.3.13) умножим на $t > 0$ и вычтем результаты из (6.3.12) и (6.3.13) соответственно. После преобразования находим

$$\sum_{i=1}^{m}(x_{li} - tx_{ij})A_{li} + tA_j = A_0, \tag{6.3.14}$$

$$\sum_{i=1}^{m}(x_{li} - tx_{ij})c_{li} + tc_{lj} = Z_0 - t(Z_j - c_j) \tag{6.3.15}$$

При условии неотрицательности коэффициентов в (6.3.14) выражение (6.3.14) определяет новое допустимое решение задачи $X_1 = (x_{li} - tx_{1j}, ..., x_{lm} - tx_{mj})$. Так как все $x_{li} > 0$, то найдется такое $t > 0$, при котором вектор X_1 будет новым допустимым решением задачи со значением функционала $Z = Z_0 - t(Z_j - c_j)$. При $t > 0$ и $Z_j - c_j > 0$ $Z < Z_0$. Итак, можно построить допустимое решение с меньшим значением целевой функции. Пусть для некоторого фиксированного j $Z_j - c_j > 0$ и по крайней мере один из коэффициентов x_{ij} будет положительным. При t выбранном по наименьшему симплексному отношению, т.е. при $t = \min\limits_{(x_{ij}>0)} \frac{x_{li}}{x_{ij}}$, получаем новое допустимое решение $\vec{X}_1$ со значением функционала $Z < Z_0$. В силу невырожденности задачи минимум симплексных отношений достигается для одного значения индекса i. Если для нового допустимого базисного решения хотя бы одна из разностей $Z_j - c_j$>0, то можно построить следующее допустимое базисное решение, более близкое к оптимальному, чем имеющееся. Очевидно, этот процесс заканчивается тогда, когда для некоторого базисного решения задачи все разности $Z_j - c_j \le 0$.

Можно показать, что если целевая функция Z неограничена на множестве L, то можно построить такое новое допустимое решение задачи, которое содержит $m+1$ положительную компоненту и которому отвечает сколь угодно малое значение целевой функции.

6.3.3. Критерий оптимальности допустимого базисного решения

Теорема. *Если для некоторого допустимого базисного решения* $\vec{X}_0 = (x_{l_1}, x_{l_2}, ..., x_{l_m})$ *все разности* $Z_j - c_j \le 0$, *то* $\vec{X}_0$ *является оптимальным решением.*

Доказательство. Пусть $Y = (y_1, y_2, ..., y_n) -$ произвольное допустимое решение задачи. Этому решению соответствуют следующие разложения вектора A_0 правых частей и значение функционала

$$A_0 = \sum_{j=1}^{n} y_j A_j, \tag{6.3.16}$$

$$Z^* = \sum_{j=1}^{n} y_j c_j, \tag{6.3.17}$$

Для доказательства теоремы достаточно показать, что $Z_0 \le Z^*$. Так как $Z_j - c_j \le 0 \;\; \forall j = \overline{1,n},$ то $Z_j \le c_j$, и, подставляя в (6.3.17) вместо c_j величины Z_j получим следующее неравенство: $Z^* \ge \sum_{j=1}^{n} y_j Z_j$. Подставим в это соотношение, вместо Z_j, его выражение (6.3.13) $Z^* \ge \sum_{j=1}^{n} y_j \sum_{i=1}^{m} c_{li} x_{ij}$.

Изменим порядок суммирования, тогда

$$Z^* \ge \sum_{i=1}^{m} (\sum_{j=1}^{n} y_j x_{ij}) c_{li}. \tag{6.3.18}$$

Подставляя в (6.3.16) вместо A_j, его разложение (6.3.12) и меняя порядок суммирования, найдем

$$A_0 = \sum_{j=1}^{n} y_j \sum_{i=1}^{m} x_{ij} A_{li} = \sum_{i=1}^{m} (\sum_{j=1}^{n} y_j x_{ij}) A_{li} \qquad (6.3.19)$$

Сравнивая два разложения (6.3.10) и (6.3.19) вектора A_0 по одному и тому же базису, в силу единствнннности разложения получим

$$\sum_{j=1}^{n} y_j x_{ij} = x_{lj}, i = \overline{1,m}.$$

Используя последнее соотношение, перепишем неравенство (6.3.18) в виде

$Z^* \geq \sum_{i=1}^{m} x_{li} c_{li} = Z_0$. Значит, $Z^* \geq Z_0$. Теорема доказана.

6.3.4. Алгоритм симплексного метода

Решение задачи линейного программирования симплексным методом начинается с известного допустимого опорного плана, которому отвечает единичный базис. Для возможности применения симплексного метода ее необходимо записать в стандартной форме. Матрица основных ограничений в стандартной форме записи содержит по крайней мере m различных единичных векторов, из которых можно составить единичную матрицу m-го порядка; каждое ее основное ограничение в стандартной форме записи содержит базисную переменную, которая с коэффициентом +1 входит только в одно это ограничение; правые части уравнений системы ограничений неотрицательны. Таким образом, если известно исходное допустимое базисное решение задачи с единичным базисом, то модель задачи имеет вид:

$$Z = \sum_{j=1}^{n} c_j x_j \to \min;$$

$$\begin{cases} x_1 + \quad \sum_{j=m+1}^{n} a_{1j} x_j = b_1 \\ \ldots \qquad \ldots \qquad \ldots \\ \qquad x_m + \sum_{j=m+1}^{n} a_{mj} x_j = b_m \end{cases} ; x_j \geq 0;\ j = \overline{1,m}.$$

Для удобства изложения вычислительной процедуры симплексного метода первоначальные данные о задаче и характеристики имеющегося базисного решения сводят в так называемую симплексную таблицу. Структура этой таблицы хорошо видна из следующего ее схематического изображения.

Записав информацию об исходной задаче в симплексную таблицу, получаем так называемую исходную или нулевую симплексную таблицу.

Симплекс-таблица.

Базис	$\vec{C}_Б$	$\overleftarrow{B}$	c_1	…	c_m	c_{m+1}	…	c_n	θ
			x_1	…	x_m	x_{m+1}	…	x_n	
x_1	c_1	b_1	1	…	0	a_{1m+1}	…	a_{1n}	
…	…	…	…	…	…	…	…	…	
x_m	c_m	b_m	0	…	1	a_{mm+1}	…	a_{mn}	
Δ_j		Δ_0	0	…	0	Δ_{m+1}	…	Δ_n	

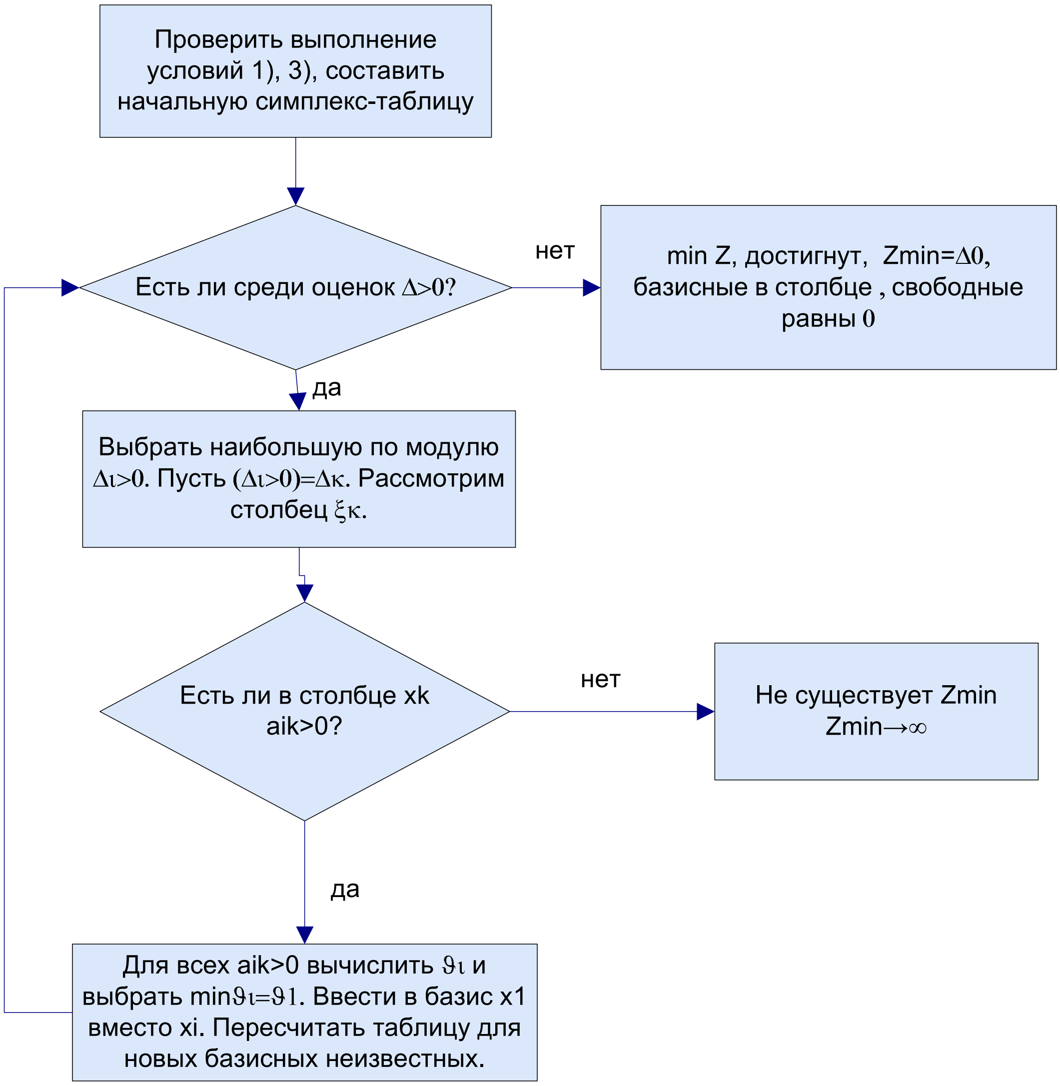
Симплекс-таблица заполняется следующим образом.

Столбцы: «Базис» – базисные неизвестные системы ограничений; «$\overline{C}_Б$» – вектор коэффициентов при базисных неизвестных в исходном линейном функционале. *Z; B* – вектор правых частей системы ограничений; «$x_1,\dots,x_n$» – матрица системы ограничений. Строки: «$c_1 \dots c_n$» – коэффициенты линейного функционала (функции цели *Z*), Δ_j – строка оценок функции *Z*, это коэффициенты при неизвестных, но при условии, что Z выражена только через свободные неизвестные, поэтому оценки в колонках под базисными неизвестными всегда равны 0. Δ_0 – значение функции цели *Z* на данном опорном плане.

$$\Delta j = \vec{C}_Б \cdot \vec{A}_j - c_j, \ \Delta_0 = \vec{C}_Б \cdot \vec{B}.$$

Таким образом, можно вычислить оценки как скалярное произведение векторов, координаты одного из которых находятся в колонке «$\vec{C}_Б$», а другого – в колонке под соответствующим неизвестным x_j ,а затем вычесть из полученного скалярного произведения величины c_j – коэффициенты в строке «$c_1,\dots,c_n$». Для вычисления значения функции на данном базисном решении можно скалярно перемножить векторы $\vec{C}_Б$ и $\vec{B}$ (координаты в соответствующих столбцах). Столбец «θ» предназначен для записи симплексных отношений $\theta = \dfrac{b_i}{a_{ik}}$ и определяется только при $a_{ik}>0$. При пересчетах таблицы симплекс-методом по наименьшему симплексному отношению выбирают неизвестное, которое выводится из базиса. Это действие соответствует анализу задачи, когда требуется увеличить одно из неизвестных так, чтобы все остальные остались неотрицательными.

Приведем алгоритм симплекс-метода для решения задачи на *min* функции Z.



Условием оптимальности плана задачи линейного программирования на min при ее решении симплекс-методом является неположительность оценок функции цели $\Delta_j \le 0,\ j=\overline{1,n}$. При решении задачи линейного программирования на max $\Delta_j \ge 0,\ j=\overline{1,n}$ условием оптимальности плана будет неотрицательность оценок.

6.4. МОДИФИЦИРОВАННЫЙ СИПЛЕКСНЫЙ МЕТОД

Анализ симплексного метода показывает, что основой перехода от одного допустимого базисного решения к другому является разложение каждого из векторов A_j, не входящих в базис, по векторам этого базиса.

Имея эти разложения, можно:

– найти разности, определяющие, какой из векторов следует ввести в базис рассматриваемого допустимого решения, или указывающие на его оптимальность;
– определить по наименьшему симплексному отношению вектор, выводящийся из базиса;
– преобразовать симплексную таблицу по формулам полного исключения с выбранным ведущим элементом и получить новое допустимое базисное решение.

Рассмотрим стандартную задачу линейного программирования $Z = CX \to \min$; $AX = A_0$; $X \geq 0$ и известно ее допустимое базисное решение с базисом $A_{l1}, A_{l2}, ..., A_{lm}$. Построим матрицу $A_{Á} = (A_{l1}, A_{l2}, ..., A_{lm})$, тогда: $A_{Á}X = A_0$, $X \geq 0$. Следовательно,

$$X = A_{Á}^{-1} A_0. \tag{6.4.1}$$

Пусть $X_j = (x_{1j}, x_{2j}, ..., x_{mj})$ – вектор A_j в данном базисе, тогда

$$X_j = A_{б}^{-1} A_j. \tag{6.4.2}$$

Вычислим

$$\Delta_j = Z_j - c_j = \sum_{i=1}^{m} c_{li} x_{ij} - c_j = c_{Á} x_j - c_j = c_{Á} A_{Á}^{-1} A_j - c_j. \tag{6.4.3}$$

Процесс вычислений закончен, если все $\Delta_j \leq 0$.

Информация, необходимая для перехода от одного допустимого базисного решения к другому, может быть получена из формул (6.4.1) – (6.4.3), если для каждого базиса известна матрица $A_{Á}^{-1}$, обратная к той, которая составлена из его векторов $A_{Á}$ и условия задачи в виде матрицы A и векторов A_0 и C. Эти соображения лежат в основе модифицированного симплексного алгоритма.

Основное различие между симплексным методом и его модифицированным вариантом состоит в том, что в первом случае по формулам полного исключения преобразовывается вся симплекс-таблица. Во втором случае по тем же самым формулам преобразовываются только элементы обратной матрицы $A_{Á}^{-1}$. В связи с этим в модифицированном симплекс-методе сокращается объем запоминаемой информации и уменьшается объем вычислений.

Пусть $\overline{X}$ – допустимое базисное неоптимальное решение и требуется перейти от базиса $A_б = (A_{l_1}, A_{l_2}, ..., A_{l_{r-1}}, A_s, A_{l_{r+1}}, ..., A_{l_m})$ к новому базису $\overline{A}_{Á} = (A_{l_1}, A_{l_2}, ..., A_{l_{r-1}}, A_k, A_{l_{r+1}}, ..., A_{l_m})$ путем замены базисного вектора $A_s = A_{lr}$ на вектор A_k. Введем обозначения

$$A_{Á}^{-1} = (a_{il_i}^{-1}),\ \overline{A}_{Á}^{-1} = (\overline{a}_{il_i}^{-1}),$$

тогда $\overline{a}_{il_i}^{-1}$ можно получить из $a_{il_i}^{-1}$ по формулам полного исключения:

$$\overline{a}_{il_i}^{-1} = a_{ij}^{-1} - \frac{a_{rj}^{-1}}{x_{rk}} x_{ik}\,(i \neq r), i = \overline{1,m},\ j = l_i;$$

$$\overline{a}_{rj}^{-1} = \frac{a_{rj}^{-1}}{x_{rk}}\,(i = r), i = \overline{1,m},\ j = l_i.$$

Вычисления в модифицированном симплексном методе проводят следующим образом. Вначале исходную задачу, записанную в стандартной форме, заменяют эквивалентной:

$$x_{n+k+1} \to \max;$$

$$\sum_{j=1}^{n} a_{ij} x_j = a_i,\ i = \overline{1,m}; \quad \sum_{j=1}^{n} c_j x_j + x_{n+k+1} = 0,$$

$$x_j \geq 0;\ j = \overline{1,n}\ (0 \leq m-k \leq m),$$

где $m-k$ – количество базисных переменных с коэффициентом +1, содержащихся в основных ограничениях задачи и не входящих в ее целевую функцию;

x_{n+k+1} – вспомогательная свободная от ограничений на знак переменная, такая, что $x_{n+k+1} = -Z = -\sum_{j=1}^{n} c_j x_j$.

Так же, как и в обычном симплексном методе, модифицированный процесс вычислений начинают с известного допустимого базисного решения с естественным или частично искусственным, или полностью искусственным единичным базисом. В базис в модифицированном симплексном методе можно брать только те единичные вектор-столбцы матрицы A, которым в целевой функции отвечают коэффициенты, равные нулю.

6.5. ДВОЙСТВЕННОСТЬ В ЛИНЕЙНОЙ ОПТИМИЗАЦИИ

Каждой задачей линейного программирования соответствует некоторая другая линейная задача, называемая двойственной или сопряженной. Рассматриваемая задача по отношению к своей двойственной называется исходной или прямой. Рассмотрим экономический смысл двойственной задачи.

В качестве исходной задачи возьмем задачу производственного планирования. Пусть в распоряжении предприятия имеются m видов ресурсов в количествах, равных $a_1, a_2, ..., a_m$ соответственно. Ресурсы должны быть использованы для производства n видов продукции, стоимость единицы которой известна и равна c_j $(j = \overline{1,n})$. Кроме того, известны нормы a_{ij} потребления каждого из ресурсов на производство единицы каждого вида продукции. План производства $\vec{X} = (x_1, x_2, ..., x_n)$ требуется составить из условия максимизации общей стоимости продукции $Z = \sum_{j=1}^{n} c_j x_j \to \max$

при ограничениях на использование ресурсов

$$\sum_{j=1}^{n} a_{ij} x_j \le b_j,\ i = \overline{1,m};$$

$$x_j \ge 0, \quad j = \overline{1,n}$$

По исходным данным этой задачи сформулируем другую экономическую задачу. Для этого предположим, что предприятию разрешено на его усмотрение реализовать все указанные ресурсы. В связи с этим возникает необходимость установить оптимальные цены $y_1, y_2,\ldots,y_m$ на эти ресурсы, пользуясь следующими соображениями: 1) покупатель ресурсов стремится минимизировать их общую стоимость; 2) с другой стороны, предприятие за каждый вид ресурсов желает выручить сумму, не меньшую той, которую оно может получить в результате использования этих ресурсов. Это объясняется тем, что в противном случае ему выгоднее организовать переработку имеющихся ресурсов. риведенная формулировка позволяет получить следующую математическую модель: минимизировать общую стоимость всех ресурсов $F = \sum_{i=1}^{m} b_i y_i \to \min$

при условиях: $$\begin{cases} \sum_{i=1}^{m} a_{ij} y_i \ge c_j, \ j = \overline{1,n}; \\ y_i \ge 0, \quad i = \overline{1,m}. \end{cases}$$

Сравнивая эти задачи, можно отметить следующее: 1) количество неизвестных в одной задаче равно числу основных ограничений другой; 2) матрица коэффициентов при неизвестных в основных ограничениях одной из задач является транспонированной по отношению к матрице ограничений второй задачи; 3) неравенства в основных ограничениях обеих задач имеют противоположный смысл; 4) коэффициенты целевой функции одной из задач являются правыми частями второй задачи; 5) в одной из задач целевая функция максимизируется, а в другой – минимизируется.

Общие правила построения двойственных задач

Исходная задача	Двойственная задача
1) $Z = \sum_{k=1}^{n} c_k x_k \to \max$	$F = \sum_{i=1}^{m} b_i\, y_i \to \min$
2) $\sum_{k=1}^{n} a_{ik} x_k \le b_i,\, i = \overline{1,s}$;	$y_i \ge 0,\, k = \overline{1,s}$;
3) $\sum_{k=1}^{n} a_{ik} x_k = b_i,\, i = \overline{s+1,m}$;	$x_i \overset{\ge}{\underset{\le}{}} 0,\, i = \overline{s+1,m}$;
4) $x_k \ge 0,\, k = \overline{1,l},\, l \le n$;	$\sum_{i=1}^{m} a_{ik} y_i \ge c_k,\, k = \overline{1,l}$;
5) $x_k \overset{\ge}{\underset{\le}{}} 0,\, k = \overline{l+1,n}$.	$\sum_{i=1}^{m} a_{ik} y_i = c_k,\, k = \overline{l+1,n}$.

1) В задачах ограничения – неравенства следует записывать со знаком «≤» при максимизации и со знаком «≥» при минимизации.

2) Каждому i–му значению исходной задачи соответствует переменная y_i двойственной задачи и, наоборот, каждому k–му ограничению двойственной задачи соответствует переменная x_k исходной задачи.

3) Каждому i–му ограничению – неравенству исходной задачи соответствует в двойственной задаче условие не отрицательности переменной ($y_i \geq 0$), а равенству – переменная без ограничений на знак (любого знака). Наоборот, неотрицательной переменной x_k соответствует в двойственной задаче k–l ограничение неравенство, а переменной произвольного знака – равенство.

4) Матрицы системы ограничений двойственной пары задач взаимно транспонированы. Следовательно, строка коэффициентов a_{ik} в k–ом ограниче-нии двойственной задаче столбец коэффициентов при x_k в ограничениях исходной задачи и наоборот.

Свободные члены ограничений одной из задач являются коэффициентами при соответствующих переменных в целевой функции другой задачи. При этом максимизация меняется на минимизацию, и наоборот.

Замечание. Соотношение двойственности взаимное, т.е. задача, двойственная по отношению к двойственной, совпадает с исходной.

В случаях, когда исходная задача задана в симметричной или канонической форме, получаем частные виды сопряженных пар задач.

Симметричная пара двойственных задач

Исходная задача	Двойственная задача
$Z = \sum_{k=1}^{n} c_k x_k \to \max$	$F = \sum_{i=1}^{m} b_i\, y_i \to \min$
$\sum_{k=1}^{n} a_{ik} x_k \leq b_i,\ i = \overline{1,m};$	$y_i \geq 0,\ k = \overline{1,m}\ ;$
$x_k \geq 0,\ k = \overline{1,n}.$	$\sum_{i=1}^{m} a_{ik} y_i \geq c_k,\ k = \overline{1,n}.;$

Несимметричная пара двойственных задач

Исходная задача

$$Z = \sum_{k=1}^{n} c_k x_k \to \max$$

$$\sum_{k=1}^{n} a_{ik} x_k = b_i,\ i = \overline{1,m};$$

$$x_k \geq 0,\ k = \overline{1,n}.$$

Двойственная задача

$$F = \sum_{i=1}^{m} b_i y_i \to \min$$

$$y_i \overset{\geq}{\underset{\leq}{-}} 0,\ k = \overline{1,m}\ ;$$

$$\sum_{i=1}^{m} a_{ik} y_i \geq c_k,\ k = \overline{1,n}.;$$

Пример. Построить двойственную задачу к заданной:

$$Z = x_1 - 2x_2 + x_3 - x_4 + x_5 \to min$$

$$\begin{cases} x_1 - 2x_2 + x_3 + 3x_4 - 2x_5 = 6 \\ 2x_1 + 3x_2 - 2x_3 - x_4 + x_5 \leq 4 \\ x_1 \qquad + 3x_3 \qquad - 4x_5 \geq 8 \end{cases} \left|\begin{matrix} y_1 \\ y_2 \\ y_3 \end{matrix}\right.$$

$$x_1 \geq 0,\ x_3 \geq 0,\ x_5 \geq 0.$$

Решение. Прежде чем приступить к построению двойственной задачи, необходимо упорядочить запись исходной задачи. Так как целевая функция минимизируется, то неравенства должны быть записаны в виде «≥». Для этого второе неравенство умножим на -1, после чего оно запишется в виде

$$-2x_1 - 3x_2 + 2x_3 + x_4 - x_5 \geq -4.$$

Введя переменные y_1, y_2 и y_3, запишем в соответствии с указанным правилом двойственную задачу:

$$F = 6y_1 - 4y_2 + 8y_3 \to max$$

$$\begin{cases} y_1 - 2y_2 + y_3 \leq 1, \\ -2y_1 - 3y_2 \qquad = -2, \\ y_1 + 2y_2 + 3y_3 \leq 1, \\ 3y_1 + y_2 \qquad = -1, \\ -2y_1 - y_2 - 4y_3 \geq 1. \end{cases} \qquad y_2 \geq 0,\ y_3 \geq 0,\ y_1 \overset{\leq}{\underset{>}{-}} 0.$$

Второе и четвертое ограничения выражены в виде равенств, так как соответствующие им переменные x_2 и x_4 не подчинены условиям неотрицательности. Условия неотрицательности в двойственной задаче наложены только на переменные y_2 и y_3 ,так как им соответствуют в исходной задаче ограничения в виде неравенств.

6.5.1. Свойства решений прямой и двойственной задач. Основные теоремы теории двойственности

Первая теорема двойственности

Если одна из задач из пары двойственных задач обладает оптимальным решением, то и другая имеет оптимальное решение, причем экстремальные значения соответствующих целевых функций равны

$$\max Z = \min F.$$

Если же у одной из этих задач целевая функция неограничена, то двойственная ей задача не имеет допустимых решений. Наконец, если одна из этих задач не имеет допустимых решений, то двойственная ей задача либо также не имеет допустимых решений, либо имеет неограниченную целевую функцию.

Пример 1. Для задачи, состоящей в определении максимального значения функции $Z = 2x_1 + 6x_2$ при условиях

$$\begin{cases} -2x_1 + 3x_2 \le 12, \\ \quad x_1 + x_2 \le 9, \end{cases}$$

$$x_1, x_2 \ge 0$$

составить двойственную задачу.

Решение. Двойственной задачей по отношению к исходной является задача, состоящая в определении минимального значения функции $F = 12\,y_1 + 9\,y_2$ при условиях

$$\begin{cases} -2y_1 + y_2 \ge 2, \\ 3\,y_1 + y_2 \ge 6, \end{cases} \qquad y_1, y_2 \ge 0$$

Как в исходной, так и в двойственной задаче число неизвестных равно двум. Следовательно, их решение можно найти, используя геометрическую интерпретацию задачи линейного программирования. Значения целевых функций исходной и двойственной задач на их оптимальных планах должны совпадать.

Пример 2. Требуется найти решение двойственной пары ЗЛП:

$Z = x_2 - x_4 - 3x_5 \rightarrow min$ $\qquad\qquad$ $F=y_1+5y_2+2y_3 \rightarrow max$

$$\begin{cases} x_1 + 2x_2 \quad - x_4 + x_5 \quad = 1 \\ \quad 3x_2 \quad\quad + x_5 + x_6 = 5 \\ -4x_2 + x_3 + 2x_4 - x_5 \quad = 2 \end{cases} \left| \begin{matrix} y_1 \\ y_2 \\ y_3 \end{matrix} \right.$$

$x_1 \geq 0, \quad j = \overline{1,6}$

$$\begin{cases} y_1 \quad\quad \leq 0 \\ 2y_1 + 3y_2 - 4y_3 \leq 1 \\ \quad\quad y_3 \leq 0 \\ -y_1 \quad 2y_3 \leq -1 \\ y_1 \quad + y_2 - \ y_3 \leq -3 \\ \quad y_2 \quad\quad \leq 0 \end{cases}$$

Приведем только начальную и последнюю симплекс-таблицы решения исходной задачи. Ее решать предпочтительнее, т.к. в ней число уравнений меньше числа неизвестных, а также она уже приведена к каноническому виду и в системе ограничений выделен базис.

Базис	$\vec{N}$	$\vec{B}$	0	1	0	-1	-3	0
			x_1	x_2	x_3	x_4	x_5	x_6
x_1	0	1	1	2	0	1	1	0
x_6	0	5	0	3	0	0	1	1
x_3	0	2	0	4	1	2	-1	0
Δ_j		0	0	-1	0	1	3	0
…………………………………………………………………………………								
x_5	-3	4	2	0	1	0	1	0
x_4	1	1/3	-2/3	1	-1/3	0	0	1/3
x_2	-1	1/3	-1/3	0	1/3	1	0	2/3
Δ_j	-46/3	-19/3	0	-11/3	0	0	0	-1/3

Проверим: $y_i \leq 0$, $i=1,2,3$, что соответствует системе ограничений двойственной задачи.

$$F_{max}=-19/3+5\cdot(-1/3)+2\cdot(-11/3)=-46/3=Z_{min}.$$

Из последней симплекс-таблицы найдем:

$y_1=\Delta_1+c_1 =-19/3 + 0 = -19/3;$

$y_2=\Delta_6 +c_6 =-1/3 + 0 = -1/3;$ $y_3=\Delta_3 +c_3 =-11/3 + 0 = -1/3;$

Ответы: исходная задача: $Z_{min} = - 46/3;$ $\vec{X}^T_{opt} = (0;11/3;\ 0;\ 1/3;\ 4;\ 0)$:

двойственная задача: $F_{max} = - 46/3;$ $\vec{Y}^T_{opt} = (-\ 19/3;\ -\ 1/3;\ -\ 11/3).$

Вторая теорема двойственности

Первая формулировка. Пусть $\vec{X} = (\bar{x}_1, \bar{x}_2, ..., \bar{x}_n)$, $\vec{Y} = (\bar{y}_1, \bar{y}_2, ..., \bar{y}_m)$ – *допустимые решения прямой и двойственной задач соответственно. Для* *того чтобы эти решения были оптимальными, необходимо и достаточно, чтобы они удовлетворяли следующему условию. Если значение какой-либо переменной допустимого решения прямой задачи отлично от нуля, то допустимое решение двойственной задачи обращает в строгое равенство соответствующее этой переменной ограничение двойственной задачи.*

Эту же теорему можно сформулировать более кратко.

Вторая формулировка. Для того, чтобы два допустимых решения $\vec{X}=(\bar{x}_1,\bar{x}_2,...,\bar{x}_n)$, $\vec{Y}=(\bar{y}_1,\bar{y}_2,...,\bar{y}_m)$ *пары двойственных задач были оптимальны, необходимо и достаточно, чтобы эти решения удовлетворяли так называемым « условиям дополняющей нежесткости»:*

$$1)\ \bar{x}_j\left(\sum_{i=1}^{m}a_{ij}\bar{y}_i-c_j\right)=0,\quad j=\overline{1,n};\qquad 2)\ \bar{y}_i\left(\sum_{i=1}^{m}a_{ij}\bar{x}_j-b_i\right)=0,\quad i=\overline{1,m};$$

Т.е. чтобы равнялось нулю произведение значения любой переменной одной задачи на разность между значениями левой и правой частей соответствующего ограничения другой задачи.

Очевидно, что для несимметричных пар достаточно выполнения только условий 1) или 2).

Пример 3. Проверить вектор $\vec{X}_0^T=(2{,}0{,}0{,}5)$ на оптимальность для следующей задачи $Z=5x_1+12x_2+8x_3+11x_4\to min;$

$$\begin{cases} x_1+2x_2+x_3+2x_4\ge 12,\\ x_1+4x_2-3x_3+8x_4\ge 6,\\ x_1+5x_2-x_3+x_4\le 7,\quad x_j\ge 0;\ j=\overline{1,4}\end{cases}$$

Построим двойственную задачу:

$Z=5x_1+12x_2+8x_3+11x_4\to min;$ $\qquad F=12y_1+6y_2\text{-}7y_3\to max$

$$\begin{cases} x_1+2x_2+x_3+2x_4\ge 12,\\ x_1+4x_2-3x_3+8x_4\ge 6,\\ -x_1-5x_2+x_3-x_4\ge -7,\quad x_j\ge 0;\ j=\overline{1,4}\end{cases}\qquad \begin{cases} y_1+y_2-y_3\le 5,\\ 2y_1+4y_2-5y_3\le 12,\\ y_1-3y_2+y_3\le 8,\\ 2y_1+8y_2-y_3\le 11,\\ x_j\ge 0;\ j=\overline{1,4}\end{cases}$$

Запишем условия дополняющей жесткости 1):

$$\begin{cases} 2\cdot\ (y_1+y_2-y_3-5)=0, \\ 0\cdot(2y_1\ +4y_2-5y_3-12)=0, \\ 0\cdot(y_1-3y_2+\ \ y_3-8)=0, \\ 5\cdot(2y_1+8y_2-\ \ y_3-11)=0, \end{cases}$$

Если произведение двух сомножителей равно нулю, то один из них строго равен нулю. Следовательно, $\begin{cases} y_1+y_2-y_3-5=0, \\ 2y_1+8y_2-\ \ y_3-11=0. \end{cases}$

Запишем условия дополняющей жесткости 2):

$\begin{cases} y_1\cdot(2+0+0+10+2)=0, \\ y_2\cdot(2+0-0+4-6)=0, \\ y_3\cdot(-2+0+0-5+7)=0. \end{cases}$ $\begin{cases} y_1\cdot 0=0, \\ y_2\cdot 36=0, \\ y_3\cdot 0=0. \end{cases}$ следовательно, $y_2=0$.

При записи условий дополнительной жесткости 2) одновременно проверяем $\vec{X}_0$ на допустимость, т.е. числа в скобках (разность левой и правой частей соответствующего ограничения исходной задачи) неотрицательны.

Таким образом, получена следующая система уравнений для определения вектора $\vec{Y}_0$, отвечающего $\vec{X}_0$:

$$\begin{cases} y_1+\ \ \ y_2-y_3=5, \\ 2y_1+8y_2-\ \ y_3=11, \\ \qquad\quad y_2\qquad\ =0. \end{cases}$$

Решив ее, получим: $y_1=6, y_2=0, y_3=1;$ $\vec{Y}_0=(6;0;1)$.

Проверим $\vec{Y}_0=(6;0;1)$ на допустимость. Очевидно, достаточно проверить, удовлетворяет ли $\vec{Y}_0$ второму и третьему основным ограничениям двойственной задачи.

Заметим, что тот же результат можно было получить. Используя 1–ю формулировку второй теоремы двойственности.

Для этого подставим $\vec{X}_0$ в систему ограничений исходной задачи.

$$\begin{cases} 2+2\times 0+0+2\times 5=12=12, \\ 2+4\times 0\text{-}3\times 0+8\times 5=42>6, \\ \text{-}2\text{-}5\times 0+0\text{-}\ \ \ 5=\text{-}7=\text{-}7. \end{cases}$$

По второй теореме двойственности $y_2=0$, т.к. для $\vec{X}_0$ второе ограничение исходной задачи выполняется как строгое неравенство (42>6), а $y_1, y_3>0$, т.к. первое и третье ограничения выполняются как равенства. По этим условиям получим ту же систему уравнений для определения $\vec{Y}_0$, что и при записи условий дополнительной нежесткости.

Проверим допустимые решения $\vec{X}_0$ и $\vec{Y}_0$ пары двойственных задач на оптимальность. Для этого в данном случае достаточно убедиться в справедливости равенства $Z(\vec{X}_0) = F(\vec{Y}_0)$ =65.

При решении задач подобного типа не только проверяется на оптимальность вектор одной из задач, но и одновременно, в случае положительного ответа на поставленный вопрос, ищется оптимальное решение двойственной задачи.

Третья теорема двойственности или теорема об оценках

Значения переменных y_i в оптимальном решении двойственной задачи представляют собой оценки влияния свободных членов ограничений исходной задачи на экстремальное значение ее целевой функции Z_{max}, т.е. $y_i = \dfrac{\partial Z_{\max}}{\partial b_i}$.

Из этой теоремы вытекает, что при малых изменениях Δb_i, правых частей основных ограничений задачи приращение целевой функции

$$\Delta Z = \vec{Y}_{opt} \cdot \Delta\vec{B} = \sum_{i-1}^{m} y_i \, \Delta b_i .$$

При более значительных изменениях правых частей ограничений выполняется следующее оценочное неравенство: $\vec{Y}_{1\,opt} \cdot \Delta\vec{B} \le \Delta Z_{\max} \le \vec{Y}_{opt} \cdot \Delta\vec{B}$

где $\vec{Y}_{1\,opt}$ – оптимальное решение двойственной задачи при измененных значениях $\vec{B} + \Delta\vec{B}$ правых частей основных ограничений исходной задачи. Если же $\vec{Y}_{1\,opt} = \vec{Y}_{opt}$, то = $\Delta Z_{max} = \vec{Y}_{opt} \cdot \Delta\vec{B}$

Пример 4. Рассмотрим пару двойственных задач.

Исходная задача

$Z=-4x_1 +2x_2-x_3+4x_4 - x_5+18 \rightarrow max$

$$\begin{cases} x_1 + 2x_2 + x_3 + 2x_4 + x_5 = 8, \\ -4x_1 + 7x_2 - x_3 - 5x_4 + 2x_5 = 7, \\ x_j \ge 0, \quad j = \overline{1,5}. \end{cases}$$

Двойственная задача

$F=8y_1+7y_2+18 \rightarrow min$

$$\begin{cases} y_1 - 4y_2 \ge -4, \\ 2y_1 + 7y_2 \ge 2, \\ y_1 - y_2 \ge -1, \quad y_i \substack{\ge \\ \le} 0. \\ 2y_1 - 5y_2 \ge 4, \\ y_1 + 2y_2 \ge -1, \end{cases}$$

Легко убедиться в том, что их оптимальными решениями будут

$\overrightarrow{X}_{opt} = \left(0, 9/4, 0, 7/4, 0\right)$, $\overrightarrow{Y}_{opt} = \left(19/12, -1/6\right)$, $Z_{\max} = F_{\min} = 59/2$.

Проверим, как изменится Z_{max} в данном примере, если свободный член $b_1 = 8$ первого уравнения исходной задачи увеличится до 10.

Можно было бы решить заново исходную задачу при новых значениях свободных членов: $b_1^* = 10$, $b_2 = 7$. В результате получили бы

$$\overrightarrow{X}_{opt} = \left(0, 8/3, 0, 7/3, 0\right), \qquad Z^*_{\max} = 98/3$$

К этому же результату можно прийти, если воспользоваться формулой $\Delta Z_{max} = \sum_{i-1}^{m} y_i \, \Delta b_i$. В данном примере

$$\Delta Z_{\max} = Z^*_{\max} - Z_{\max} = y_1 \cdot \Delta b_1 + y_2 \cdot \Delta b_2 = 19/2\,(b_1^* - b_1) - 1/6 \cdot 0 = 19/6.$$

Отсюда $Z^*_{\max} = Z_{\max} + 19/6 = 98/3.$

6.5.2. Экономическая интерпретация основной и двойственной задач

Основную задачу линейного программирования можно экономически интерпретировать следующим образом. Пусть для производства некоторого продукта можно использовать ***n*** технологий. При этом используются ***m*** различных видов сырья и прочих производственных факторов (ингредиентов). По j-й технологии в единицу времени расходуется a_{ij} единиц j-го ингредиента и производится c_j единиц продукта. Пусть по j-й технологии производство ведется x_j единиц времени. Требуется отыскать такой план $\overrightarrow{X} = (x_1, x_2, ..., x_n)$ при котором из имеющихся запасов ингредиентов выпускалось бы максимальное количество продукта. Математическая модель сформулированной задачи такова:

$$Z = \sum_{j=1}^{n} c_j x_j \to \max;$$

$$\sum_{j=1}^{n} a_{ij} x_j \le b_j, \; i = \overline{1, m};$$

$$x_j \ge 0, \quad j = \overline{1, n}.$$

Для экономической интерпретации двойственной задачи в качестве двойственных переменных y_i, $(i=\overline{1,m})$ возьмем стоимость единицы j-го ингредиента, использовав в качестве масштабной единицы стоимость единицы выпускаемого продукта. В двойственной задаче требуется найти такой вектор $\vec{Y}=(y_1,y_2,...,y_m)$ цен на имеющиеся виды ингредиентов, при котором суммарная стоимость запасов ингредиентов, подлежащих использованию, минимальна: $F=\sum_{i=1}^{m} a_i y_i \to \min$ при непревышении стоимости продукта, произведенного по каждой технологии, стоимости затраченных по этой технологии ингредиентов:

$$\sum_{i=1}^{m} a_{ij} y_i \geq c_j, \quad j=\overline{1,n}$$

Рентабельность плана $F=\sum_{i=1}^{m} a_i y_i \to \min$ должна означать точное воплощение в стоимости произведенного по этому плану продукта всей стоимости запаса ингредиентов (полное отсутствие непроизводительных затрат), т.е. в рентабельном плане стоимость всех затрат производства должна равняться стоимости произведенного продукта $\sum_{i=1}^{m} \mathrm{b_i}\, y_i = \sum_{j=}^{n} c_j x_j$ т.е. $F=Z$.

Обозначим через $\overrightarrow{X_0}=(x_1^0,x_2^0,...,x_n^0)$ и $\overrightarrow{Y_0}=(y_1^0,y_2^0,...,y_m^0)$ оптимальные решения сформулированной пары двойственных задач. Тогда по второй основной теореме двойственности для этой пары оптимальных решений должны выполняться условия дополняющей нежесткости:

$$1)\ x_j^0\left(\sum_{i=1}^{m} a_{ij} y_i^0 - c_j\right)=0, \quad j=\overline{1,n}; \qquad 2)\ y_i^0\left(\sum_{i=1}^{m} a_{ij} x_j^0 - b_i\right)=0, \quad i=\overline{1,m};$$

Эти условия для данной задачи имеют важное экономическое значение. Действительно, если j-я технология не рентабельна. Т.е. затраты по j-й технологии превышают стоимость произведенного по этой технологии продукта $\sum_{i=1}^{m} a_{ij} y_i^0 > c_j,\ x_j^0=0$. Это означает, что в оптимальном плане время работы по нерентабельной технологии равно нулю.

Если же *j*-я технология применяется в оптимальном плане, т.е. $x_j^0 > 0$, то по первому условию обязательно $\sum_{i=1}^{m} a_{ij} y_i^0 - \mathrm{c_j} = 0.$

Это означает, что стоимость затраченных по этой технологии средств равна стоимости произведенного продукта. Другими словами, если в оптимальном плане $\mathrm{x_j^0} > 0$, то соответствующая технология рентабельна. Таким образом, оптимальные значения двойственных переменных являются инструментом оценки эффективности (рентабельности) технологий (производственно–технологических способов).

Пусть $\mathrm{y_i^0} > 0$, тогда из условий 2) следует, что *i*-й ингредиент используется полностью: $\sum_{j=1}^{n} a_{ij} x_j^0 - \mathrm{bi} = 0,$ т.е. *i*-й ингредиент является дефицитным ресурсом.

Если *i*-й ингредиент используется не полностью (не дефицитен): $\sum_{j=1}^{n} a_{ij} x_j^0 < \mathrm{bi},$ то из условий 2) следует, что оптимальная двойственная оценка такого ресурса равна нулю: $\mathrm{y_i^0} = 0$

Таким образом, оптимальные двойственные оценки являются мерой дефицитности ресурсов.

Проведенный анализ показывает, что оптимальные планы рассмотренной пары двойственных задач тесно взаимосвязаны в том смысле, что оптимальному плану производства отвечает оптимальный план относительных цен ингредиентов. По оптимальному плану относительных цен ингредиентов все применяемые технологии рентабельны, а неприменяемые – убыточны. При этом ингредиенты с положительной стоимостью используются полностью, а неиспользуемые полностью – имеют нулевую стоимость.

РАЗДЕЛ 7. ЦЕЛОЧИСЛЕННЫЕ ЗАДАЧИ ЛИНЕЙНОГО ПРОГРАММИРОВАНИЯ

В значительной части экономических задач, относящихся к задачам линейного программирования, компоненты решения должны выражаться в целых числах, т.е. быть целочисленными. К ним относятся, задачи, в которых переменные означают количество единиц неделимой продукции, например: число станков при загрузке оборудования, число вычислительных машин в управляющем комплексе, задачи определения вариантов мощностей предприятий, технологических способов производства продукции, распределения транспортных средств по рейсам и многие другие.

Численные методы решения задач целочисленного линейного программирования можно условно разбить на три группы: **методы отсечения, комбинаторные и приближенные.**

7.1. МЕТОД ОТСЕЧЕНИЯ

Среди *методов отсечения* - широко известный *метод Гомори.* Основная идея метода Гомори заключается в следующем. Вначале задача линейного программирования решается без ограничений целочисленности. Если полученное решение удовлетворяет ограничениям целочисленности, то оно является оптимальным решением задачи целочисленного линейного программирования.

Если же все или часть компонентов оптимального плана принимают дробные значения, то к ограничениям исходной задачи добавляется линейное ограничение, которому удовлетворяют все целочисленные решения исходной задачи, но не удовлетворяет полученное нецелочисленное решение. Затем решается полученная расширенная задача, т.е. выполняется одна итерация. Если полученное решение не удовлетворяет условиям целочисленности, то добавляется еще одно ограничение и т.д. Описанная процедура отсечения продолжается конечное число шагов до получения целочисленного оптимального решения либо до выявления неразрешимости задачи в целых числах.

Задача линейного целочисленного программирования формулируется следующим образом: *найти такое решение (план $X=(x_1, x_2, \ldots x_n)$), при котором линейная функция*

$$Z = \sum_{j=1}^{n} c_j x_j \tag{7.1.1}$$

принимает максимальное или минимальное значение при ограничениях

$$\sum_{i=1}^{n} a_{ij} x_j = b_i, \quad i = \overline{1,m}. \qquad x_j \geq 0, \quad j = \overline{1,n}. \tag{7.1.2}$$

$$x_j \text{ – целые числа.} \tag{7.1.3}$$

Следует отметить, что классическая транспортная задача и некоторые другие задачи транспортного типа «автоматически» обеспечивает решение задачи в целых числах (если, конечно, целочисленны параметры условий). Однако в общем случае условие целочисленности (7.1.3), добавляемое к обычным задачам линейного программирования, существенно усложняет ее решение.

Для решения задач линейного целочисленного программирования используется ряд методов. Самый простой из них обычный метод линейного программирования. В случае если компоненты оптимального решения оказываются нецелочисленными, их округляют до ближайших целых чисел. Этот метод применяют тогда, когда отдельная единица совокупности составляет малую часть объема всей совокупности. В противном случае округление может привести к далекому от оптимального целочисленному решению, поэтому используют специально разработанные методы.

Остановимся подробнее на методах отсечения.

7.1.1. Метод Гомори

Сущность методов отсечения состоит в том, что сначала задача решается без условия целочисленности. Если полученный план целочисленный, задача решена. В противном случае к ограничениям задачи добавляется новое ограничение, обладающее следующими свойствами:

1) *оно должно быть линейным;*

2) *должно отсекать найденный оптимальный нецелочисленный план:*

3) *не должно отсекать ни одного целочисленного плана.*

Дополнительное ограничение, обладающее указанными свойствами, называется *правильным отсечением.*

Далее задача решается с учетом нового ограничения. После этого в случае необходимости добавляется еще одно ограничение и т.д.

Геометрически добавление каждого линейного ограничения отвечает проведению прямой (гиперплоскости), которая отсекает от многогранника (многоугольника) решений некоторую его часть вместе с оптимальной точкой с нецелыми координатами, но не отсекает ни одной из целых точек этого многогранника. В результате новый многогранник решений содержит все целые точки, заключавшиеся в первоначальном многограннике решений и, соответственно полученное оптимальное решение будет целочисленным.

Один из алгоритмов решения задачи линейного целочисленного программирования (7.1.1)–(7.1.3), предложенный Гомори, основан на симплексном методе и использует достаточно простой способ построения правильного отсечения.

Пусть задача линейного программирования (7.1.1)–(7.1.3) имеет конечный оптимум и на последнем шаге ее решения симплексным методом получены следующие уравнения, выражающие базисные переменные $x_1, x_2, \dots, x_j, \dots, x_m$ через свободные переменные $x_{m+1}, x_{m+2}, \dots, x_{m+i}, \dots, x_n$ оптимального решения

$$\begin{cases} x_1 = \beta_1 - \alpha_{1m+1}x_{m+1} - \cdots - \alpha_{1n}x_n, \\ x_2 = \beta_2 - \alpha_{2m+1}x_{m+1} - \cdots - \alpha_{2n}x_n, \\ \dots \ \dots \ \dots \ \dots \ \dots \ \dots \ \dots \ \dots, \\ x_i = \beta_i - \alpha_{im+1}x_{m+1} - \cdots - \alpha_{in}x_n, \\ \dots \ \dots \ \dots \ \dots \ \dots \ \dots \ \dots \ \dots, \\ x_m = \beta_m - \alpha_{mm+1}x_{m+1} - \cdots - \alpha_{mn}x_n, \end{cases} \tag{7.1.4}$$

так, что оптимальным решением задачи (7.1.1)–(7.1.2) является $X^* = (\beta_1, \beta_2, \dots, \beta_m, 0, 0, \dots, 0)$, в котором, например, β_I – нецелая компонента. В этом случае можно доказать, что условие [♥]

$$\{\beta_i\} - \{\alpha_{im+1}\}x_{m+1} - \dots - \{\alpha_{in}\}x_n \le 0, \tag{7.1.5}$$

[♥] В неравенстве (7.1.5) присутствует символ {}, означающий дробную часть числа. *Целой частью числа а* называется наибольшее целое число *[a]*, не превосходящее *a*, а дробной частью числа – число *{a}*, равное разности между этим числом и его целой частью, т.е. *{a}=a-[a]*.

сформулированное по i-му уравнению системы (7.1.4), обладает всеми свойствами правильного отсечения.

Для решения задачи целочисленного линейного программирования (7.1.1)–(7.1.3) методом Гомори используется следующий *алгоритм*:

1. Симплексным методом решить задачу (7.1.1)–(7.1.2) без учета условия целочисленности. Если все компоненты оптимального плана целые, то он является оптимальным и для задачи целочисленного программирования (7.1.1)–(7.1.2). Если первая задача (7.1.1)–(7.1.2) неразрешима (т.е. не имеет конечного оптимума или условия ее противоречивы), то вторая задача (7.1.1)–(7.1.3) также неразрешима.

2. Если среди компонент оптимального решения есть нецелые, то выбрать компоненту с наибольшей целой частью и по соответствующему уравнению системы (7.1.3) сформировать правильное отсечение (7.1.5).

3. Неравенство (7.1.5) введением дополнительной неотрицательной целочисленной переменной преобразовать в равносильное уравнение

$$\{\beta_i\}-\{\alpha_{im+1}\}x_{m+1}-\ldots-\{\alpha_{in}\}x_n+x_{n+1}=0, \qquad (7.1.6)$$

и включить его в систему ограничений (7.1.2).

4. Полученную расширенную задачу решить симплексным методом. Если найденный оптимальный план будет целочисленным, то задача целочисленного программирования (7.1.1)–(7.1.3) решена. В противном случае вернуться к п. 2 алгоритма.

Если задача разрешима в целых числах, то после конечного числа шагов (итераций) оптимальный целочисленный план будет найден.

Если в процессе решения появится уравнение (выражающее базисную переменную через свободные) с нецелым свободным членом и целыми остальными коэффициентами, то соответствующее уравнение не имеет решения в целых числах. В этом случае данная задача не имеет целочисленного оптимального решения.

Пример. Для приобретения оборудования по изготовлению цемента фирма выделяет 34 тыс. грн. Оборудование должно быть размещено на площади, не превышающей 60 кв. м. Фирма может заказать оборудование двух видов: менее мощные машины типа А стоимостью 3 тыс. грн, требующие производственную площадь 3 кв.м и обеспечивающие производительность за смену 2 т цемента, и более мощные машины типа *B* стоимостью 4 тыс. грн, занимающие площадь 5 кв. м и обеспечивающие производительность за смену 3 т цемента.

Требуется составить оптимальный план приобретения оборудования, обеспечивающий максимальную общую производительность при условии, что фирма может приобрести не более 8 машин типа В.

Решение. Обозначим через x_1, x_2 количество машин типа *A* и *B* соответственно, через *Z* – общую производительность. Тогда математическая модель задачи примет вид: $Z=2x_1+3x_2 \to max$

при ограничениях: $\begin{cases} 3x_1+5x_2 \le 60, & (1) \\ 3x_1+4x_2 \le 34, & (2) \\ \qquad x_2 \le 8, & (3) \\ x_1 \ge 0, \quad x_2 \ge 0 \end{cases}$ x_1, x_2– целые числа.

Приведём задачу к каноническому виду, введя дополнительные неотрицательные переменные x_3, x_4, x_5. Получим систему ограничений.

$$\begin{cases} 3x_1+5x_2+x_3=60, & (1) \\ 3x_1+4x_2+x_4=34, & (2) \\ \qquad x_2+x_5=8, & (3) \\ x_j \ge 0, \quad j=\overline{1,5}. \end{cases}$$

Решаем задачу симплексным методом.

Баз	$\vec{C}_Б$	$\vec{B}$	2 x_1	3 x_2	0 x_3	0 x_4	0 x_5	θ
x_3	0	60	3	5	1	0	0	20
x_4	0	34	3	4	0	1	0	34/3
x_5	0	8	0	1	0	0	1	8-min
Δ_j		0	-2	-3	0	0	0	
x_3	0	20	3	0	1	0	-5	20/3
x_4	0	2	3	0	0	1	-4	2/3
x_2	3	8	0	1	0	0	1	
Δ_j		24	-2	0	0	0	3	
x_3	0	18	0	0	1	-1	-1	
x_1	2	2/3	1	0	0	1/3	-4/3	
x_2	3	8	0	1	0	0	1	
Δ_j		76/3	0	0	0	2/3	1/3	

Получено оптимальное решение: X_{opt} $=(2/3;8;18;0;0)$; Z_{max} $=76/3$ $=25\frac{1}{3}$, которое не удовлетворяет условиям целочисленности. Выразим базисные переменные через свободные:

$$\begin{cases} x_1 = \frac{2}{3} - \frac{1}{3}x_4 + \frac{4}{3}x_5, \\ x_2 = 8 \quad - \quad x_5, \\ x_3 = 18 + x_4 \quad + x_5 \end{cases}$$

Функция цели $Z = 25\frac{1}{3} - \frac{2}{3}x_4 - \frac{1}{3}x_5$. По первому уравнению с переменной x_1, получившей нецелочисленное значение в оптимальном решении (2/3), составляем дополнительное ограничение: $\left\{\frac{2}{3}\right\} - \left\{\frac{1}{3}\right\}x_4 - \left\{-\frac{4}{3}\right\}x_5 \le 0$.

Обращаем внимание на то, что мы берем дробную часть *свободного члена с тем же знаком*, который он имеет в уравнении, а дробные части *коэффициентов при не основных переменных x_4 и x_5 – с противоположными знаками.*

Так как дробные части

$$\left\{\frac{2}{3}\right\} = \left\{0 + \frac{2}{3}\right\} = \frac{2}{3}, \quad \left\{\frac{1}{3}\right\} = \left\{0 + \frac{1}{3}\right\} = \frac{1}{3}, \qquad \left\{-\frac{4}{3}\right\} = \left\{-2 + \frac{2}{3}\right\} = \frac{2}{3},$$

то последнее неравенство запишем в виде

$$\frac{2}{3} - \frac{1}{3}x_4 - \frac{2}{3}x_5 \le 0.$$

Введя дополнительную целочисленную переменную x_6 ≥ 0, получим равносильное последнему неравенству уравнение $\frac{2}{3} - \frac{1}{3}x_4 - \frac{2}{3}x_5 + x_6 = 0$, или для введения в симплекс-таблицу: $-\frac{1}{3}x_4 - \frac{2}{3}x_5 + x_6 = -\frac{2}{3}$.

Вводим дополнительное последнее уравнение в систему, полученную на последнем шаге решения задачи (без условия целочисленности). Тогда симплекс-таблица примет вид:

Баз	$\vec{C}_Б$	$\vec{B}$	2 x_1	3 x_2	0 x_3	0 x_4	0 x_5	0 x_6
x_3	0	18	0	0	1	-1	-1	0
x_1	2	2/3	1	0	0	1/3	-4/3	0
x_2	3	8	0	1	0	0	1	0
x_6	0	-2/3	0	0	0	-1/3	-2/3	1
Δ_j		76/3	0	0	0	2/3	1/3	0
x_3	0	19	0	0	1	-1/2	0	-3/2
x_1	2	2	1	0	0	1	0	-2
x_2	3	7	0	1	0	-1/2	0	3/2
x_5	0	1	0	0	0	1/2	1	-3/2
Δ_j		25	0	0	0	1/2	0	1/2

Базисное решение $X_4 = \left(\frac{2}{3};8;18;0;0;-\frac{2}{3}\right)$ недопустимое. (Заметим, что после включения в систему ограничений дополнительного уравнения, соответствующего правильному отсечению, всегда будет получаться недопустимое базисное решение).

Для получения допустимого базисного решения необходимо перевести в основные переменную, входящую с положительным коэффициентом в уравнение, в котором свободный член отрицательный, т.е. x_1 или x_5 (на этом этапе линейную функцию не рассматриваем). Переводим в основные, например, переменную x_5.

$$\overline{X}_{opt} = (2;7;19;0;1;0); \quad Z_{max} = 25$$

— получено оптимальное решение, удовлетворяющее условиям целочисленности.

Максимальную производительность 25т цемента за смену можно получить приобретением 2 машин типа *А* и 7 машин типа *В*; при этом незанятая площадь помещения составит 19 кв. м, остатки денежных средств из выделенных равны *0*, в резерве для покупки – 1 машина типа *В* (шестая компонента содержательного смысла не имеет).

7.2. ТРАНСПОРТНАЯ ЗАДАЧА

7.2.1. Определение транспортной модели

Транспортная задача занимает особое место среди задач линейного программирования, специфическая структура которой позволяет разработать эффективные методы, основанные на теории двойственности. Частными случаями транспортной задачи является задача о назначениях и транспортная задача с промежуточными пунктами.

Следует отметить, что классическая транспортная задача и некоторые другие задачи транспортного типа «автоматически» обеспечивают решение задачи в целых числах (если, конечно, целочисленные параметры условий).

Транспортные модели описывают перемещение (перевозку) какого-либо товара из пункта отправления в пункт назначения. Цель транспортной задачи – определение объёмов перевозок из пунктов отправления в пункты назначения с минимальной суммарной стоимостью перевозок. При этом должны учитываться ограничения, накладываемые на объёмы грузов, имеющих в пунктах отправления (предложения), и ограничения, учитывающие потребность грузов в пункты назначения (спрос). В транспортных задачах предполагается, что стоимость перевозки по какому-либо маршруту прямо пропорциональна объёму груза, перевозимого по этому маршруту. В общем случае транспортную модель можно применить для описания ситуаций, связанных с управлением движения капиталов, составлением расписания, назначением персонала.

На Рис.7.1 показано представление транспортной задачи в виде сети с m пунктами отправления и n пунктами назначения, которые показаны в виде узлов сети. Дуги, соединяющие узлы сети, соответствуют маршрутам, связывающим пункты отправления и назначения. С дугой (i,j), соединяющей пункт отправления i с пунктом назначения j, соотносятся два вида данных: 1) c_{ij} – стоимость перевозки единицы груза из пункта i с пунктом назначения j и 2) x_{ij} – количество перевозимого груза. Объём грузов в пункте отправления i равен a_i, а объём грузов в пункте назначения j равен b_j. Задача состоит в определении неизвестных величин x_{ij}, минимизирующих суммарные транспортные расходы и удовлетворяющие ограничениям, накладываемым на объёмы грузов в пунктах отправления и назначения.

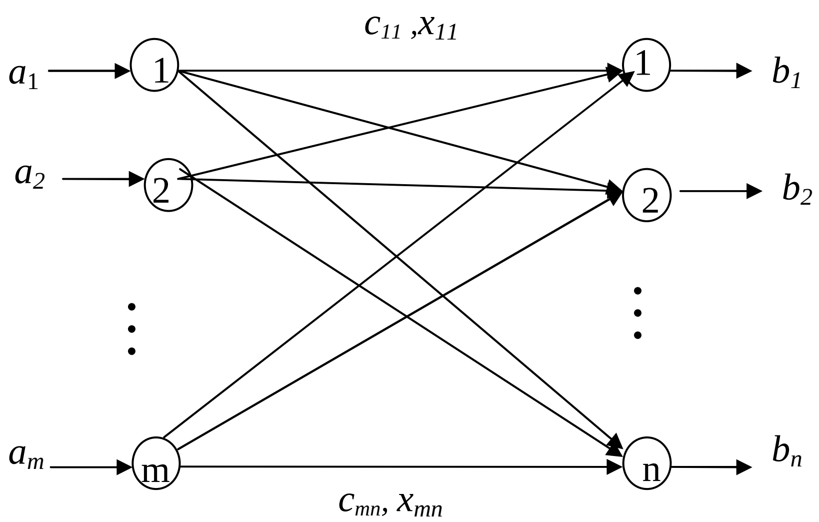


Рис. 7.1

7.2.2. Математическая модель задачи

Пусть A_i, $i=\overline{1,m}$ – поставщики грузов, у каждого из них имеется a_i – условных единиц однородного груза. Этот груз требуется доставить потребителям B_j, $j=\overline{1,n}$, причем каждому из них необходимо b_j условных единиц груза. Перевозки надо осуществить так, чтобы их стоимость была минимальной.

Чтобы составить математическую модель задачи, введем неизвестные, которые составят план перевозок: пусть x_{ij} $(i=\overline{1,m};\ j=\overline{1,n})$ – количество условных единиц груза, который потребуется перевезти от i-того поставщика j-тому потребителю.

Будем считать известными тарифы перевозок, т.е. c_{ij} – стоимость перевозки единицы данного груза от i-того поставщика j-тому потребителю ($i=\overline{1,m}$; $j=\overline{1,n}$). Тогда общая стоимость перевозок, которую требуется минимизировать:

$$Z = c_{11}x_{11} + c_{12}x_{12} + \ldots + c_{mn}x_{mn} = \sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}x_{ij}\ .$$

Z – функция цели данной задачи линейного программирования.

Получим теперь ограничения транспортной задачи. Очевидно, суммарное количество условных единиц груза, отправленного от любого поставщика A_i, $i=\overline{1,m}$ должно быть равно имеющемуся у него количеству, т.е.:

$$x_{i1} + x_{i2} + \ldots + x_{in} = a_i;\ (i=\overline{1,m}) \text{ или } \sum_{j=1}^{n} x_{ij} = a_i;\ (i=\overline{1,m});$$

Полученные *m* уравнений называются ограничениями по запасам или горизонтальными уравнениями.

С другой стороны, каждому потребителю должно быть доставлено строго требуемое количество условных единиц груза, следовательно

$$x_{1j} + x_{2j} + \ldots + x_{mj} = b_j;\ (j = \overline{1,n}) \quad \text{или} \quad \sum_{i=1}^{m} x_{ij} = b_j;\ (j = \overline{1,n}).$$

Полученные *n* уравнений называются ограничениями по потребностям.

Если свести все данные и искомые величины в таблицу, то получим так называемую транспортную таблицу (табл. 19.2)

Таблица 7.1

Поставщик	Потребитель				Запасы груза
	B₁	**B₂**	**B₃**	**B₄**	
A_1	c_{11} x_{11}	c_{12} x_{12}	…	c_{1n} x_{1n}	a_1
A_2	c_{21} x_{21}	c_{22} x_{22}	…	c_{2n} x_{2n}	a_2
…	…	…	…	…	…
A_m	c_{m1} x_{m1}	c_{m2} x_{m2}	…	c_{mn} x_{mn}	a_m
Потребность	b_1	b_2	…	b_n	

Горизонтальные уравнения системы ограничений могут быть получены суммированием перевозок (x_{ij}) по строкам таблицы, а вертикальные – по столбцам. В правом верхнем углу каждой клетки таблицы перевозок проставляются тарифы (стоимость) перевозок.

Таким образом, приходим к следующей задаче линейного программирования: найти такое неотрицательное решение системы ограничений

$$\begin{cases} \sum_{j=1}^{n} x_{ij} = a_i;\ (i = \overline{1,m}), \\ \sum_{i=1}^{m} x_{ij} = b_j;\ (j = \overline{1,n}), \end{cases}$$

чтобы оно давало минимум линейной функции стоимости перевозок:

$$Z = \sum_{i=1}^{m} \sum_{j=n} c_{ij} x_{ij} \to \min,$$

Транспортная задача всегда решается в предположении, что суммарная потребность в грузах равна суммарному запасу, т.е.: $\sum_{i=1}^{m} a_i = \sum_{j=1}^{n} b_j$

Теорема. *Любая транспортная задача, у которой суммарный объем запасов совпадает с суммарным объемом потребностей, имеет решение.*

Такая задача называется "закрытой" транспортной задачей. Если это не так, то задача называется "открытой" и перед решением приводится к "закрытой" путем введения условного потребителя B_{n+1} с объёмом потребностей $b_{n+1} = \sum_{i=1}^{m} a_i - \sum_{j=1}^{n} b_j$ (если $\sum_{i=1}^{m} a_i > \sum_{j=1}^{n} b_j$), или условного поставщика A_{m+1} с объёмом запасов a_{m+1} = $\sum_{j=1}^{n} b_j - \sum_{i=1}^{m} a_i$ а, (если $\sum_{i=1}^{m} a_i < \sum_{j=1}^{n} b_j$), стоимости перевозок в этих пунктах считаются равными нулю.

7.2.3. Алгоритм решения транспортной задачи

Алгоритм решения транспортной задачи повторяет основные шаги симплекс-метода. Однако для представления данных, вместо обычных симплекс-таблиц, используются транспортные таблицы со специальной структурой (табл. 7.1). Последовательность шагов алгоритма решения транспортной задачи в точности повторяет аналогичную последовательность этапов симплексного алгоритма.

Шаг 1. Определим *начальное* базисное допустимое решение (опорный план). Если каким-либо способом получен невырожденный опорный план транспортной задачи, то в матрице (x_{ij}) ($i = \overline{1,m}$; $j = \overline{1,n}$) значений его компонент положительными являются только $m+n-1$, а остальные равны нулю.

Клетки, в которых находятся отличные от нуля перевозки, называются *занятыми,* остальные – *свободными.* Занятые клетки соответствуют базисным неизвестным, и для невырожденного опорного плана их количество равно $m+n-1$. Опорность плана при записи условий транспортной задачи в виде таблицы заключается в его ацикличности, т.е. в таблице нельзя построить замкнутый цикл, все вершины которого лежат в занятых клетках.

Шаг 2. На основании оптимальности симплекс-метода среди небазисных переменных определяем *вводимую в базис*. Если все небазисные переменные удовлетворяют условию оптимальности, вычисления заканчиваются; в противном случае переходим к третьему шагу.

Шаг 3. С помощью условия допустимости симплекс-метода среди текущих базисных переменных определяем исключаемую. Затем находим новое базисное решение. Возвращаемся ко второму шагу.

Рассмотрим каждый описанный шаг в отдельности.

7.2.4.Определение начального опорного плана.

Рассмотрим два метода построения начального опорного плана перевозок.

I метод – метод северо-западного угла. Он не учитывает стоимость перевозки единицы груза (тариф), поэтому полученные опорные планы далеки от оптимальных. Но поскольку алгоритм этого метода очень прост, то его удобно применять при компьютерных расчетах.

Пример. Пусть условия транспортной задачи записаны в таблице 7.2.

Таблица 7.2.

оставщик	Потребитель				Запасы груза
	B1	**B2**	**B3**	**B4**	
A1	5 **200**	1 **0**	4 **0**	3 **0**	200
A2	4 **50**	2 **250**	8 **0**	1 **0**	300
A3	6 **0**	2 **200**	1 **200**	4 **0**	400
A_4	2 **0**	4 **0**	6 **300**	3 **200**	500
Потребность	250	450	500	200	

Проверим задачу на "закрытость" – подсчитаем суммарные запасы груза: $\sum_{i=1}^{m} a_i$ =1400. ед. груза и суммарные потребности $\sum_{j=1}^{n} b_j$ = 1400. ед. груза. Следовательно, задача является "закрытой" и не требует введения условных поставщиков или потребителей. Начиная с ее левого верхнего (северо-западного) угла начинаем удовлетворение потребностей в грузе у всех потребителей последовательно. Первый поставщик, имея 200 ед. груза, отправляет первому потребителю 200 ед. Вычеркиваем клетки первой строки, т.к. у первого поставщика нет больше груза. Недостающие 50 ед. груза 1-й потребитель берёт у второго поставщика. Вычеркиваем оставшиеся клетки первого столбца, т.к. 1-й потребитель полностью удовлетворён. Второй поставщик оставшиеся 250 ед. груза поставляет второму потребителю, у которого в результате оказывается 250 ед. груза, что меньше его потребности. Вычеркиваем оставшиеся клетки второй строки. Третий поставщик из имеющихся у него 400 ед. груза поставляет 200 ед. второму потребителю – тогда его потребность полностью удовлетворена и оставшуюся во 2-ом столбце клетку можно вычеркнуть. Оставшиеся 200 ед. третий поставщик отдает 3-му потребителю. Наконец четвертый поставщик поставляет 300 ед. груза третьему потребителю и 200 ед. – четвертому. Весь груз распределен, все потребности удовлетворены. Стоимость перевозок на этом плане:

$$Z=200\cdot5+50\cdot4+250\cdot2+200\cdot2+200\cdot1+300\cdot6+200\cdot3=4700 \text{ усл. ден. ед.}$$

План является невырожденным, т.к. содержит $m+n-1= =4+4-1=7$ базисных клеток.

7.2.5. II метод - метод минимальной стоимости

Суть метода заключается в том, что из всей таблицы стоимостей выбирают наименьшую стоимость и в клетку, которая ей соответствует, помещают меньшее из чисел a_i или bj. Затем из рассмотрения исключают либо строку, соответствующую поставщику, запасы которого полностью израсходованы, либо столбец, соответствующий потребителю, потребности которого полностью удовлетворены, либо и строку и столбец, если израсходованы запас поставщика и удовлетворены потребности потребителя. Из оставшейся части таблицы стоимостей снова выбирают наименьшую стоимость, и процесс распределения запасов продолжают, пока все запасы не будут распределены, а потребности удовлетворены.

Составим с помощью этого метода опорный план уже рассмотренной задачи.

Таблица 7.3.

	Поставщик	Потребитель				Запасы груза
u_j		**B1**	**B2**	**B3**	**B4**	
u_1	**A1**	5 **0**	1 **200**	4 **0**	3 **0**	**200**
u_2	**A2**	4 **0**	2 **100**	8 **0**	1 **200**	**300**
u_3	**A3**	6 **0**	2 **0**	1 **400**	4 **0**	**400**
u_4	**A_4**	2 **250**	4 **150**	6 **100**	3 **0**	**500**
	Потребности	**250**	**450**	**500**	**200**	
	v_i	v_1	v_2	v_3	v_4	

Найдем клетку с наименьшей стоимостью перевозок и поставим в нее наибольшую возможную перевозку. В данном случае $c_{12}=c_{24}=c_{33}=1$, но $x_{12}=200$, $x_{24}=300$, а $x_{33}=400$, т.е. наибольшая из возможных перевозок $x_{33}=400$. Теперь поставщик A_3 полностью исчерпал свои запасы и в дальнейшем в 3-ю строку ничего больше не заносим (заштриховываем). Находим следующую клетку с минимальной стоимостью, $c_{24}=1$ и заносим в нее $x_{24}=200$. Теперь 4-й потребитель полностью удовлетворён (заштриховываем). Далее в клетку c_{12} поставляем наиболее возможную перевозку – 200. Теперь поставщик A_1 полностью израсходовал запасы (заштриховываем 1-ю строку).

Аналогично $c_{22}=c_{32}=c_{41}=2$. Но $x_{32}=0$, т.к. поставщик A_3 исчерпал свои запасы, а $x_{22}=100$ (т.к. потребитель B_2 уже получил *200 усл. ед*. груза от поставщика A_1, то берём у поставщика A_2 недостающие *100 усл. ед.* груза). Поэтому заполняем $x_{41}=250$. Первый потребитель получил груз полностью, поэтому заштриховываем первый столбец.

Среди оставшихся незаполненных и не заштрихованных клеток выбираем клетку с наименьшей стоимостью: $c_{42}=4$. Поставим в нее перевозку $x_{42}=150$. Теперь потребитель B_2 получил весь требующийся груз, поэтому заштриховываем 2-й столбец. Затем в клетку с $c_{43}=6$ должны поставить перевозку $x_{43}=100$ и тогда полностью будут исчерпаны запасы четвёртого поставщика (заштриховываем 4-ю строку). Просуммировав перевозки по горизонтальным и вертикальным строкам таблицы и получив величины соответственно запасов и потребностей, проверим, что система ограничений удовлетворяется. Перевозки, которые отвечают заполенным клеткам, называются базисными, остальные, значения которых равны нулю, свободными. Стоимость перевозок при данном распределении перевозок:

$$Z=250\cdot 2+200\cdot 1+100\cdot 2+150\cdot 4+400\cdot 1+100\cdot 6+200\cdot 1=2700 \text{ усл. ден. ед.}$$

Очевидно, что этим методом получен начальный опорный план с существенно меньшей стоимостью перевозок, следовательно, гораздо более близкий к оптимальному.

7.3. РЕШЕНИЕ ТРАНСПОРТНОЙ ЗАДАЧИ МЕТОДОМ ПОТЕНЦИАЛОВ

Проверка плана на оптимальность осуществляется методом потенциалов. Пусть каждому поставщику соответствует потенциал u_i $(i=\overline{1,m})$, а потребителю – v_j $(j=\overline{1,n})$. Эти величины связаны между собой соотношением $\boxed{u_i+v_j=c_{ij}}$, где пары индексов i,j соответствуют базисным клеткам, c_{ij} – стоимость перевозок в базисных клетках. В свободных клетках должно выполняться неравенство $\boxed{u_i+v_j\le c_{ij}}$. *Если во всех клетках таблицы (базисных и свободных) вышеприведенные соотношения выполняются, то план является оптимальным.* Если нет – то требуется улучшение плана, которое осуществляется переброской грузов по циклу.

Проверим методом потенциалов на оптимальность полученный опорный план (табл.7.3):

Этап 1

Очевидно, что число составленных уравнений будет на одно меньше, чем число неизвестных потенциалов.

Поэтому обычно задают один из потенциалов равным нулю, а затем подсчитывают остальные.

Полагая потенциал $u_1=0$, определяем остальные потенциалы u_i из соотношения $u_j+v_i=c_{i,j}\,(i=\overline{1,m}\,,\ j=\overline{1,n})$, просматривая все занятые клетки.

Этап 2

Определяем значения оценок $S_{i,j}=c_{i,j}-(v_j+u_i)$ для всех свободных клеток, используя найденные на 1-ом этапе значения потенциалов. Для пустых клеток условие оптимальности оценки: $S_{i,j} \geq 0$. Просматривая пустые клетки ($x_{i,j} = 0$), получим:

Занятые клетки	*Свободные клетки*
$u_1=0$,	$S_{1,1} = c_{1,1} - (v_1 + u_1) = 6$,
$v_2=c_{1,2}-u_1=1$,	$S_{1,3} = c_{1,3} - (v_3 + u_1) = 1$,
$u_2=c_{2,2}-v_2=1$,	$S_{1,4} = c_{1,4} - (v_4 + u_1) = 3$,
$u_4=c_{4,2}-v_2=3$,	$S_{2,1} = c_{2,1} - (v_1 + u_2) = 4$,
$v_4=c_{2,4}-u_2=0$,	$S_{2,3} = c_{2,3} - (v_3 + u_2) = 4$,
$v_1=c_{4,1}\ u_4=-1$,	$S_{3,1} = c_{3,1} - (v_1 + u_3) = 9$,
$v_3=c_{4,3}-u_4=3$,	$S_{3,2} = c_{3,2} - (v_2 + u_3) = 3$,
$u_3=c_{3,3}-v_3=-2$.	$S_{3,4} = c_{3,4} - (v_4 + u_3) = 6$,
	$S_{4,4} = c_{4,4} - (v_4 + u_4) = 0$.

Все условия оптимальности выполнены, следовательно, полученный план является оптимальным.

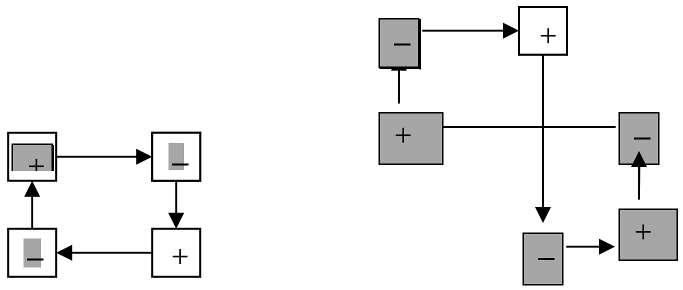
Транспортная задача решена. *Целевая функция* $Z=2700$.

Если условия оптимальности плана не выполняются, требуется улучшение плана.

7.3.1. Улучшение плана способом переброски груза по циклу

Одним из способов улучшения опорного плана – переброска груза по циклу.

Цикл – это многозвенная ломаная линия, построенная в таблице, причем она начинается и заканчивается в свободной клетке, в которой не выполняется условие оптимальности. Если таких клеток несколько, то выбирается та, в которой разность $|u_i + v_j - c_{ij}|$ – наибольшая. Каждое звено цикла соединяет две базисные клетки одной строки или одного столбца, т.е. ломаная поворачивает только под прямым углом. Простейший цикл – прямоугольный, однако возможны циклы с самопересечениями.



Очевидно, что в любом случае цикл соединяет четное число клеток. Направление движения по циклу произвольно – по часовой стрелке или против.

Переброска груза осуществляется так. Каждой клетке цикла присваивают знаки (+) или (-) чередуя их, начиная с (+) в свободной клетке, в которой не выполняется условие оптимальности. Затем выбирают наименьшую перевозку из базисных клеток со знаком (-) и прибавляют ее в клетках со знаком (+) и вычитаются в клетках со знаком (-). Например:

После переброски имеем:

Во всех клетках цикла остаются неотрицательные перевозки и сохраняется баланс (т.е. суммарные перевозки по строкам и столбцам сохраняются).

После переброски груза по циклу следует снова проверить план на оптимальность.

Пример. Четыре строительных предприятия используют цемент, который находится в четырёх пунктах. Потребности в цементе каждого из предприятий соответственно равны: 120т., 150т., 130т., 200т. Запасы в каждом из четырёх пунктов отправления соответственно равны: 160т.,140т., 170т., 130т.

Тарифы перевозок заданы матрицей $C = \begin{pmatrix} 7 & 8 & 1 & 2 \\ 4 & 5 & 9 & 8 \\ 9 & 2 & 3 & 6 \\ 1 & 4 & 5 & 3 \end{pmatrix}$. Составить такой план перевозок, при котором общая стоимость перевозок будет минимальна.

Исходные данные запишем в виде таблицы (Табл.7.4)

Таблица 7.4.

Поставщик	Потребитель				Запасы груза
	B1	**B2**	**B3**	**B4**	
A1	7	3	1	2	160
A2	4	5	2	1	140
A3	2	2	3	6	170
A4	1	4	4	3	130
Потребность	120	150	130	200	

Транспортная задача имеет закрытый тип, так как суммарный запас груза равен суммарным потребностям 160+140+170+130=120+150+130+200=600 .

Находим опорный план методом минимальной стоимости. Находим незанятую клетку с минимальным тарифом: (1,3). Помещаем туда меньшее из чисел $a_1*=160$, $b_3*=130$. Находим незанятую клетку с минимальным тарифом: (2,4). Помещаем туда меньшее из чисел $a_2^*=140$, $b_4^*=200$ и т.д.

Полученный опорный план представлен в таблице 7.5.

Таблица 7.5.

	Поставщик	Потребитель				Запасы груза
u_j		**B1**	**B2**	**B3**	**B4**	
u_1	**A1**	7	3	1 **130**	2 **30**	160
$u2$	**A2**	4	5	2	1 **140**	140
u_3	**A3**	2	2 **150**	3	6 **20**	170
u_4	**A4**	1 **120**	4	4	3 **10**	130
	Потребности	120	150	130	200	
	v_i	v_1	v_2	v_3	v_4	

Целевая функция $Z=130\cdot1+30\cdot2+140\cdot1+150\cdot2+29\cdot6+120\cdot1+10\cdot3=900.$

Проверим план на оптимальность методом потенциалов:

Этап 1. Полагая потенциал $u_1=0$, определяем остальные потенциалы из соотношения $u_j+v_i=c_{i,j}\,(i=\overline{1,4},\ j=\overline{1,4})$, просматривая все занятые клетки.

Определяем значения оценок $S_{i,j}=c_{i,j}-(v_j+u_i)$ для всех свободных клеток:

Занятые клетки	*Свободные клетки*
$u_1=0$	$S_{1,1} = c_{1,1} - (v_1 + u_1) = 7$,
$v_3=c_{1,3}-u_1= 1$,	$S_{1,2} = c_{1,2} - (v_2 + u_1) = 5$,
$v_4=c_{1,4}-u_1= 2$,	$S_{2,1} = c_{2,1} - (v_1 + u_2) = 5$,
$u_2=c_{4,2}-v_4= -1$,	$S_{2,2} = c_{2,2} - (v_2 + u_2) = 8$,
$u_3=c_{4,3}-v_4= 4$,	$S_{2,3} = c_{2,3} - (v_3 + u_2) = 2$,
$u_4=c_{4,4}-v_4= 1$,	$S_{3,1} = c_{3,1} - (v_1 + u_3) = \boxed{-2}$,
$v_2=c_{3,2}-u_3= -2$,	$S_{3,3} = c_{3,3} - (v_3 + u_3) = \boxed{-2}$,
$v_1=c_{4,1}-u_4= 0$.	$S_{4,2} = c_{4,2} - (v_2 + u_4) = 5$,
	$S_{4,3} = c_{4,3} - (v_3 + u_4) = 2$.

Оценки $S_{i,j}$ для всех клеток, удовлетворяющих условию: $x_{i,j} = 0$.

Так как имеется две клетки с одним и тем же наиболее неоптимальным значением оценки равной -2, то из них выбирается клетка, имеющая наименьший тариф. Наиболее потенциальной является клетка **(**3,1**)**.

Строим для этой клетки цикл, помечая клетки цикла знаками "плюс" и "минус" (табл.7.6).

Таблица 7.6

Поставщик	Потребитель				Запасы груза
	B1	**B2**	**B3**	**B4**	
A1	7	3	1 **130**	2 **30**	160
A2	4	5	2	1 **140**	140
A3	+ 2	2 **150**	3	- 6 **20**	170
A4	- 1 **120**	4	4	+ 3 **10**	130
Потребность	120	150	130	200	

Перемещаем по циклу груз величиной в 20 единиц, прибавляя эту величину к грузу в клетках со знаком "+" и отнимая её от груза в клетках со знаком "-". В результате перемещения по циклу получим новый план (табл.7.7).

Таблица 7.7

Поставщик	Потребитель				Запасы груза	u_i
	B1	B2	B3	B4		
A1	7 7	3 3	1 130	2 30	160	*0*
A2	4 5	5 6	2 2	1 140	140	*-1*
A3	2 20	2 150	3 0	6 2	170	*2*
A4	1 100	4 3	4 2	3 30	130	*1*
Потребность	120	150	130	200		
v_j	*0*	*0*	*1*	*2*		

Целевая функция $Z= 130\cdot1+30\cdot2+140\cdot1+20\cdot2+150\cdot2+100\cdot1+30\cdot3=860$. Значение целевой функции уменьшилось на 40 единиц по сравнению с предыдущим этапом.

Этап 2

Проверим полученный план на оптимальность. Полагая потенциал $u_1=0$ и просматривая все занятые клетки, определяем остальные потенциалы из соотношения $u_j+v_i=c_{i,j}$ $(i=\overline{1,4},\ j=\overline{1,4})$. Значения соответствующих оценок приведены в левом нижнем углу свободных клеток таблицы (табл. 7.7). Так как все условия оптимальности выполнены, то полученный план является оптимальным.

Задача. Четыре строительных участка потребляют щебень, вырабатываемый тремя дробильными установками. Производительность (в сутки) дробильных установок соответственно: 65т; 85т; 60т; 40 т. Суточная потребность в щебне на участках: 120т; 80т; 100т. Матрица стоимостей перевозок 1т щебня:

$$C=\begin{pmatrix}1&2&3\\4&5&6\\2&3&4\\3&4&5\end{pmatrix}$$

1) Составить план перевозок, минимизирующий их стоимость, при условии, что недостающее количество щебня можно не обеспечивать.

2) Решить задачу при условии, что участки должны обеспечиваться щебнем в полном объеме. Для этого сравнить два варианта: увеличить производительность установки №3 или увеличить производительность установки №4.

Решение.

Так как суммарные потребности строительных участков превышают возможности поставщиков на 50 т., то не все участки могут быть удовлетворены и данная задача является задачей открытого типа.

1) Ответим на первый вопрос задачи. Составим исходную таблицу 7.8.

Таблица 7.8

Поставщик	Потребитель			Запасы груза
	B1	**B2**	**B3**	
A1	1	2	3	65
A2	4	5	6	85
A3	2	3	4	60
A4	3	4	5	40
Потребность	120	80	100	

Введём фиктивного поставщика A_5 с запасами 50т. (Табл.7.9).

Таблица 7.9

Поставщик	Потребитель			Запасы груза
	B1	**B2**	**B3**	
A1	1	2	3	65
A2	4	5	6	85
A3	2	3	4	60
A4	3	4	5	40
A5	0	0	0	50
Потребность	120	80	100	

Находим опорный план для задачи.

Введем некоторые обозначения:
A_i^* – излишек нераспределенного груза от поставщика A_i
B_j^* – недостача в поставке груза потребителю B_j

Находим незанятую клетку с минимальным тарифом: (5,1). Помещаем туда меньшее из чисел $A_5*=50,\ B_1*=120$. Находим незанятую клетку с минимальным тарифом: (1,1). Помещаем туда меньшее из чисел $A_1^*=65,\ B_1^*=70$ Находим незанятую клетку с минимальным тарифом: (3,1). Помещаем туда меньшее из чисел $A_3^*=60$, $B_1^*=5$ и т.д. (Табл.7.10).

Таблица 7.10

Поставщик	Потребитель			Запасы груза
	B1	**B2**	**B3**	
A1	1 **65**	2 0	3 0	65
A2	4 0	5 0	6 **85**	85
A3	2 5	3 **55**	4 0	60
A4	3 0	4 **25**	5 **15**	40
A5	0 **50**	0 -1	0 -2	50
Потребность	120	80	100	

Целевая функция F=925.

Решаем задачу методом потенциалов:

Этап 1

Полагая потенциал $u_1=0$, определяем остальные потенциалы из соотношения $u_j+v_i=c_{i,j}\,(i=\overline{1,5},\ j=\overline{1,3})$, просматривая все занятые клетки (Табл. 7.1.10).

Определяем значения оценок $s_{i,j}=c_{i,j}-(v_j+u_i)$ для всех свободных клеток:

Для случая $x_{i,j}=0$ условие оптимальности оценки $s_{i,j}$ определяется следующим образом: $s_{i,j}\geq 0$. Оценки $s_{i,j}$ для свободных клеток приведены в левом нижнем углу таблицы 7.10.

Наиболее потенциальной является клетка *(5,3)*. Для нее оценка равна $\boxed{-2}$.

Строим для этой клетки цикл, помечая клетки цикла знаками "плюс" и "минус" (Табл. 7.11).

Таблица 7.11

Поставщик	Потребитель			Запасы груза
	B1	**B2**	**B3**	
A1	1 **65**	2	3	65
A2	4	5	6 **85**	85
A3	2 **+ 5**	3 **- 55**	4	60
A4	3	4 **+ 25**	5 **- 15**	40
A5	0 **- 50**	0	0 **+**	50
Потребность	120	80	100	

Перемещаем по циклу груз величиной в 15 единиц, прибавляя эту величину к грузу в клетках со знаком "плюс" и вычитая её от груза в клетках со знаком "минус". В результате перемещения по циклу получим новый план (табл. 7.12).

Таблица 7.12

Поставщик	Потребитель			Запасы груза	u_i
	B1	**B2**	**B3**		
A1	1 **65**	2 0	3 2	65	*0*
A2	4 $\boxed{-2}$	5 -2	6 **85**	85	*5*
A3	2 **20**	3 **40**	4 2	60	*1*
A4	3 0	4 **40**	5 2	40	*2*
A5	0 **35**	0 -1	0 **15**	50	*-1*
Потребность	120	80	100		
v_j	*1*	*2*	*1*		

Целевая функция F= 895.

Значение целевой функции уменьшилось на 30 единиц по сравнению с предыдущим этапом.

Этап 2. Проверим план на оптимальность.Полагая потенциал $u_1=0$, определяем остальные потенциалы из соотно-шения $u_j+v_i=c_{i,j}$ $(i=\overline{1,5}, j=\overline{1,3})$, просматривая все занятые клетки. Определяем значения оценок $s_{i,j}=c_{i,j}-(v_j-u_i)$ для всех свободных клеток (табл.7.13). Наиболее потенциальной является клетка **(2,1)**. Для нее оценка равна **-2**. Строим для этой клетки цикл, помечая клетки цикла знаками "+" и "-".

Таблица 7.13

Поставщик	Потребитель			Запасы груза
	B1	**B2**	**B3**	
A1	1 **65**	2	3	65
A2	**+** 4	5	**-** 6 **85**	85
A3	2 **20**	3 **40**	4	60
A4	3	4 **40**	5	40
A5	**-** 0 **35**	0	**+** 0 **15**	50
Потребность	120	80	100	

Перемещаем по циклу груз величиной в 35 единиц, прибавляя эту величину к грузу в клетках со знаком "+" и отнимая её от груза в клетках со знаком "-". В результате перемещения по циклу получим новый план:

Таблица 7.14

Поставщик	Потребитель			Запасы груза
	B1	**B2**	**B3**	
A1	1 **65**	2	3	65
A2	4 **35**	5	6 **50**	85
A3	2 **20**	3 **40**	4	60
A4	3	4 **40**	5	40
A5	0	0	0 **50**	50
Потребность	120	80	100	

Целевая функция F=825.

Значение целевой функции уменьшилось на 70 единиц по сравнению с предыдущим этапом.

Этап 3. Проверив план на оптимальность, получим (табл. 7.15).

Так как все условия оптимальности выполнены, т.е. для свободных клеток все $s_{i,j} \geq 0$, то полученный план является оптимальным (табл. 7.15).

Таблица 7.15.

Поставщик	Потребитель			Запасы груза
	B1	**B2**	**B3**	
A1	1 **65**	2 $\boxed{0}$	3 $\boxed{0}$	65
A2	4 **35**	5 $\boxed{0}$	6 **50**	85
A3	2 **20**	3 **40**	4 $\boxed{0}$	60
A4	3 $\boxed{0}$	4 **40**	5 $\boxed{0}$	40
A5	0 $\boxed{2}$	0 $\boxed{1}$	0 **50**	50
Потребность	120	80	100	

Целевая функция F= 825

Итак, 50 единиц груза потребителю B_3 не поставлено.

2) **Ответим на 2-й вопрос задачи.**

а) Увеличим производительность щебня на установке 3 на 50 т., тогда исходная таблица (7.16) будет иметь вид.

Таблица 7.16

Поставщик	Потребитель			Запасы груза
	B1	**B2**	**B3**	
A1	1 **0**	2 **0**	3 **0**	65
A2	4 **0**	5 **0**	6 **0**	85
A3	2 **0**	3 **0**	4 **0**	110
A4	3 **0**	4 **0**	5 **0**	40
Потребность	120	80	100	

Транспортная задача имеет закрытый тип, так как суммарный запас груза равен суммарным потребностям: 120+80+100=65+85+110+40=300.

Находим опорный план по методу минимальной стоимости (табл.7. 17).

Таблица 7.17

Поставщик	Потребитель			Запасы груза
	B1	**B2**	**B3**	
A1	1 **65**	2	3	65
A2	4	5	6 **85**	85
A3	2 **55**	3 **55**	4	110
A4	3	4 **25**	5 **15**	40
Потребность	120	80	100	

Целевая функция F=1025

С помощью метода потенциалов убедимся, что план оптимальный.

Транспортная задача решена (табл.7.17).

b) Увеличим производительность щебня на установке 4 на 50т., тогда исходная таблица будет иметь вид (табл.7.18).

Таблица 7.18.

Поставщик	Потребитель			Запасы груза
	B1	**B2**	**B3**	
A1	1 **0**	2 **0**	3 **0**	65
A2	4 **0**	5 **0**	6 **0**	85
A3	2 **0**	3 **0**	4 □	60
A4	3 **0**	4 **0**	5 **0**	90
Потребность	120	80	100	

Транспортная задача имеет закрытый тип, так как суммарный запас груза равен суммарным потребностям.

Находим опорный план задачи по методу минимальной стоимости (табл. 7.19).

Таблица 7.19

Поставщик	Потребитель			Запасы груза
	В1	**В2**	**В3**	
А1	1 **65**	2	3	65
А2	4	5	6 **85**	85
А3	2 **55**	3 **5**	4	60
А4	3	4 **75**	5 **15**	90
Потребность	120	80	100	

Целевая функция F=1075

Методом потенциалов убеждаемся, что план оптимальный.

Вывод: Целесообразнее увеличить производительность 3-й установки на 50т. (целевая функция F=1025 т.), т.к. увеличение производительности 4-й установки на 50т. дает увеличение целевой функции F=1075 т.

7.4. ЗАДАЧА О НАЗНАЧЕНИЯХ

7.4.1.Формулировка задачи (проблема выбора)

В этой задаче необходимо назначить работников на определённые работы при условии, что каждый работник может выполнять любую работу, хотя с различной степенью мастерства. Если на некоторую работу назначается работник именно той квалификации, которая необходима для её выполнения, тогда стоимость выполнения работы будет ниже, чем при назначении на данную работу работника неподходящей квалификации. Цель задачи – найти оптимальное (минимальной стоимости) распределение работников по всем заявленным работам.

Таблица 7.20

$$
\begin{array}{c|cccc|c}
 & 1 & 2 & \dots & n & \\
\hline
1 & c_{11} & c_{12} & \dots & c_{1n} & 1 \\
2 & c_{21} & c_{22} & \dots & c_{2n} & 1 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
n & c_{n1} & c_{n1} & \dots & c_{nm} & n \\
\hline
 & 1 & 1 & \dots & 1 &
\end{array}
$$

Коэффициент c_{ij} равен стоимости назначения работника i на работу j $\left(i=\overline{1,n}\right)$, $\left(j=\overline{1,n}\right)$. То, что количество работников равно количеству работ, не является ограничением общности, поскольку всегда можно ввести в модель фиктивных работников или фиктивные работы.

Задача о назначениях является частным случаем транспортной задачи, в которой работники соответствуют пунктам отправления, а работы – пунктам назначения. В данном случае все величины спроса и предложения равны 1.

Пусть необходимо выполнить n $\left(i=\overline{1,n}\right)$ работ. Для этого используются n $\left(j=\overline{1,n}\right)$ исполнителей, каждый из которых в состоянии выполнять любую работу. Известны затраты (c_{ij}) на выполнение j-м исполнителем i-й работы (табл. 7.20). Требуется назначить каждого исполнителя на одну работу так, чтобы минимизировать суммарные затраты. Чтобы сформулировать задачу о назначениях математически, введем переменную величину x_{ij}, отражающую факт назначения:

$$x_{ij}=\begin{cases}1, \textit{ если на } i-\textit{ю работу назначен } j-\textit{й исполнитель};\\ 0, \textit{ в противном случае}\end{cases}$$

Задача примет вид

$$\min: Z=\sum_{i=1}^{n}\sum_{j=1}^{n}c_{ij}x_{ij} \tag{7.4.1}$$

при ограничениях:

на каждую работу должен быть назначен только один исполнитель

$$\sum_{j=1}^{n}x_{ij}=1, \quad i=\overline{1,n}, \tag{7.4.2}$$

каждый исполнитель должен быть назначен для выполнения только одной работы

$$\sum_{i=1}^{n}x_{ij}=1, \quad j=\overline{1,n}, \tag{7.4.3}$$

условие $x_{ij}\in\{0,1\}$ называется условием булевости и может быть записано так:

$$x_{ij}=x_{ij}^2 \tag{7.4.4}$$

Прежде чем переходить к алгоритму решения задачи о назначениях, отметим, что она относится к транспортной при условии $a_i = b_j = 1$. Как и в транспортной, можно говорить об открытой задаче о назначениях, если число работ отлично от числа исполнителей. Если число работ меньше числа исполнителей *(m<n)*, вводят *n-m* фиктивных работ. Считается, что с назначением на фиктивные работы исполнителей не связаны затраты, т. е. соответствующие коэффициенты матрицы потерь равны нулю. В случае же *m>n* вводят *m-n* фиктивных исполнителей. Соответствующие элементы c_{ij} матрицы потерь можно полагать очень большими *(M)*.

Если задача о назначениях ставится при условии получения максимума эффективности, то ее сводят к задаче на минимум. Пусть дана матрица эффективностей $C=||c_{ij}||$. В каждом столбце найдем максималь-ный элемент l_j, т. е. $l_j = \max_i c_{ij}$. Построим матрицу $C' = \|c'_{ij}\| = \|l_j - c_{ij}\|$. Задача

$$\max_{X\in\Omega} : Z = \sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij} x_{ij}$$

где Ω – область допустимых решений (7.4.2 – 7.4.4), эквивалентна задаче

$$\min_{X\in\Omega} : Z' = \sum_{i=1}^{n}\sum_{j=1}^{n} (l_j - c_{ij}) x_{ij} \ .$$

Действительно: $$Z' = \sum_{i=1}^{n}\sum_{j=1}^{n} (l_j - c_{ij}) x_{ij} = \sum_{j}^{n} l_j - Z$$

Очевидно, *Z'* достигает минимума при условии, что *Z* достигает максимума, $X \in \Omega$.

Остановимся еще на одной детали. Если по каким-либо причинам запрещается выполнение какой-либо работы каким-либо исполнителем, то в матрицу потерь вводят очень большую стоимость *M*.

Задачу о назначениях можно эффективно решить точно так же, как и транспортную. Вместе с тем тот факт, что все величины спроса и предложения равны 1, привёл к разработке упрощённого алгоритма решения, названного венгерским методом. Хотя этот метод не имеет никакого отношения к транспортной задаче, он, как и метод потенциалов, основан на симплекс-методе.

7.4.2. Венгерский метод решения задачи о назначениях основан на двух простых утверждениях.

1. Решение задачи не изменится, если к любому столбцу или строке прибавить или вычесть некоторую компоненту, т. е. если план *X** – оптимальный план задачи (7.4.1)–(7.4.4), то он также оптимален для функции цели *Z'* с

матрицей $C' = \left\| c'_{ij} \right\|$, где $c'_{ij} = c_{ij} \pm u_i \pm v_j$, $i = \overline{1,n}$, $j = \overline{1,n}$, $X \in \Omega$.

2. Если все $c_{ij} \geq 0$ и найден план X^*, такой, что $\sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} x_{ij}^* = 0$, то X^*– оптимальный план.

Алгоритм метода. Идея алгоритма основана на сформулированных выше утверждениях. Путем прибавления определенным образом найденных чисел к некоторым столбцам и вычитания их из некоторых строк находят систему так называемых *независимых нулей*. Набор нулей называется независимым, если никакие два (или больше) нуля не лежат на одной линии. Если число независимых нулей равно n, то, приняв соответствующие им переменные x_{ij} равными единице, а все остальные – равными нулю, согласно утверждению (2), получим оптимальный план назначения.

Алгоритм метода состоит из предварительного шага и не более чем $(n\text{-}2)$ последовательно повторяющихся итераций.

На предварительном этапе в случае решения задачи на *max* ее преобразуют в эквивалентную задачу минимизации. На этом же этапе выделяется система независимых нулей. Каждая последующая итерация направлена на увеличение хотя бы на единицу числа независимых нулей. Как только число независимых нулей k станет равным размерности матрицы $n \times n$ – задача решена. Оптимальный план назначения определится положением независи-мых нулей на последней итерации.

Подготовительный этап. Пусть исходная задача имеет вид: $\max\limits_{X \in \Omega} : Z = \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} x_{ij}$. Находим $l_j = max \{c_{ij}\}$. Заменяем задачу максимизации задачей минимизации: $\min : Z' = \sum_{i=1}^{n} \sum_{j=1}^{n} c'_{ij} x_{ij}$, где $ñ'_{ij} = l_j - c_{ij}$. Далее в каждой строке матрицы $C' = \left\| c'_{ij} \right\|$ находим минимальный элемент, который вычитаем из всех элементов соответствующей строки матрицы. Переходим к матрице $C'' = \left\| c''_{ij} - \min\limits_{j} c'_{ij} \right\| = \left\| c''_{ij} \right\|$. Такое преобразование, согласно утверждению (1), не меняет решения исходной задачи. В результате преобразования в каждой строке матрицы C'' появляется хотя бы один нуль.

Подготовительный этап заканчивается выделением системы независимых нулей. Для этого в первом столбце матрицы выделяется произвольный нуль и отмечается звездочкой (*). Затем рассматриваются нули второго столбца. Если среди них есть такие, которые находятся не в одной строке с нулем первого столбца, уже отмеченным звездочкой, то один из них отмечается звездочкой. Аналогично рассматриваются нули для других столбцов. Отмеченные звездочками нули независимы, так как в каждой строке и в каждом столбце может быть не более одного нуля со звездочкой. Если число независимых нулей равно n, то, согласно утверждению (2), задача решена.

Общий шаг. Целью каждого последующего шага (итерации) является увеличение числа независимых нулей. Перед началом итерации выделяются знаком (+) столбцы матрицы, содержащие нули со звездочкой. Так как $k<n$, то не все столбцы окажутся выделенными. Устанавливаем, имеется ли среди невыделенных элементов матрицы хотя бы один нуль. Возможны два случая: а) среди невыделенных элементов имеется хотя бы один нуль; б) среди невыделенных элементов нулей нет. В первом случае проверяем, содержит ли строка с невыделенным нулем также нуль со звездочкой. Если да, то невыделенный нуль отмечается штрихом (0'), содержащая этот нуль строка отмечается справа знаком (+) и снимается знак выделения (+) над столбцом, в котором расположен нуль со звездочкой, лежащий в только что выделенной строке. Снятие знака выделения производится знаком $\oplus$. Если в строке с невыделенным нулем нулей со звездочкой нет, то невыделенный нуль также отмечается штрихом (0').

Во втором случае среди невыделенных элементов матрицы выбираем минимальный. Он вычитается из элементов, расположенных в невыделенных строках, и прибавляется к элементам, лежащим в выделенных столбцах. При этом новая матрица, согласно утверждению (1), эквивалентна исходной по отношению к оптимальному назначению.

Итерация оканчивается, если в строке с невыделенным нулем нет нуля со звездочкой. Невыделенный нуль, как указывалось выше, отмечается штрихом (0') и строится цепочка элементов по правилу: 1) движение начинается от нуля со штрихом (0') по столбцу к нулю со звездочкой (0*); 2) от нуля со звездочкой (0*) по строке переходим к нулю со штрихом (0'); 3) вновь от 0* к 0' по столбцу и т.д. Начало и конец цепочки – нули со штрихами. Отметим некоторые свойства построенной цепочки.

1. Цепочка по указанному правилу строится однозначно. В самом деле, в каждом столбце может находиться лишь один нуль со звездочкой 0* (эти нули независимы), а в каждой строке содержится не более одного нуля со штрихом 0'. Последнее обстоятельство связано с тем, что после того, как нуль отмечается штрихом, содержащая его строка выделяется и поэтому никакие другие ее нули не могут быть отмечены штрихом.
2. Покажем, что, начиная от 0', получим цепочку, которая обязательно заканчивается 0', т. е. в строке с 0* непременно найдется 0'. Предположим, что существует цепочка

$$\left\{c_{i_1 j_1}, c_{i_2 j_1}, c_{i_2 j_2}, ..., c_{i_n j_n}, c_{i_{n+1} j_n}\right\}, \tag{7.4.5}$$

3. где $c_{i_1 j_1}$, а $c_{i_{n+1} j_n}$. Причем цепочка не может быть продолжена. Элементы $c_{i_n j_n} = 0'$, и $c_{i_{n+1} j_n} = 0^*$ лежат в одном столбце. Так как $c_{i_{n+1} j_n} = 0^*$, то $c_{i_n j_n} = 0'$. Но это значит, что столбец j_n не выделен. С другой стороны в этом столбце имеется 0*, а именно $c_{i_{n+1} j_n} = 0^*$. Следовательно, знак выделения (+) столбца j_n был уничтожен, это указывает на наличие в i_{n+1}-й строке – 0'. Отсюда вывод: цепочка (5) может быть продолжена. Это противоречит тому, что цепочка (5) не может быть продолжена.
4. Так как в начале и конце цепочки стоит 0', то число элементов цепочки нечетно, причем 0' находится на нечетных местах.
5. Если случится, что в одном столбце с исходным 0' нет нуля со звездочкой, то цепочка вырождена и состоит из одного исходного элемента.

После построения цепочки звездочки над нулями уничтожаются, а штрихи заменяются звездочками. Все остальные звездочки над нулями, не находящимися на цепочке, также сохраняются. Так как в цепочке нулей со штрихами на один больше, чем нулей со звездочками, то в результате общее число нулей со звездочками увеличивается на единицу. На этом одна итерация заканчивается. Подсчитываем число нулей со звездочкой (независимых). Если $k=n$, то найден оптимальный план назначения. Если $k<n$, то переходим к следующей итерации.

Проиллюстрируем венгерский метод решения проблемы выбора на числовом примере.

Пример. Пусть надо решить задачу о назначениях на *max*,

$$\max : Z = \sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}x_{ij}\ \text{, где: } C = \left\| c_{ij} \right\| = \begin{pmatrix} 10 & 7 & 6 & 9 \\ 3 & 7 & 5 & 1 \\ 5 & 6 & 8 & 3 \\ 6 & 11 & 7 & 5 \end{pmatrix}$$

Подготовительный этап. Отыскиваем максимальный элемент в каждом из столбцов матрицы и вычитаем каждый элемент матрицы C из максимального элемента соответствующего столбца. В полученной матрице C′ отыскиваем минимальный элемент в каждой из строк и вычитаем их из соответствующих элементов матрицы C′:

$$1.\begin{array}{c} \begin{pmatrix} 10 & 7 & 6 & 9 \\ 3 & 7 & 5 & 1 \\ 5 & 6 & 8 & 3 \\ 6 & 11 & 7 & 5 \end{pmatrix} \\ \hline 10 \quad 11 \quad 8 \quad 9 \end{array} \xrightarrow{C'} 2.\begin{pmatrix} 0 & 4 & 2 & 0 \\ 7 & 4 & 3 & 8 \\ 5 & 5 & 0 & 6 \\ 4 & 0 & 1 & 4 \end{pmatrix}\begin{array}{|c} 0 \\ 3 \\ 0 \\ 0 \end{array} \xrightarrow{C_0} 3.\begin{pmatrix} 0^* & 4 & 2 & 0 \\ 4 & 1 & 0^* & 5 \\ 5 & 5 & 0 & 6 \\ 4 & 0^* & 1 & 4 \end{pmatrix}$$

Образуем первоначальную систему независимых нулей, отмечая их звёздочкой 0^*. Имеем матрицу C_0. Число независимых нулей $k=3$. Так как $k<n=4$, то переходим к основному повторяющемуся шагу.

Первая итерация. В матрице 4 над столбцами, содержащими нули со звёздочкой, ставим знаки выделения (+). Находим невыделенный столбец $j=4$. Ищем в невыделенном столбце нулевой элемент. Это $c_{14}=0$.

$$4.\begin{array}{c} \begin{array}{cccc} + & + & + & \end{array} \\ \begin{pmatrix} 0^* & 4 & 2 & 0 \\ 4 & 1 & 0^* & 5 \\ 5 & 5 & 0 & 6 \\ 4 & 0^* & 1 & 4 \end{pmatrix} \end{array} \quad 5.\begin{array}{cc} \begin{array}{cccc} \oplus & + & + & \end{array} & \\ \begin{pmatrix} 0^* & 4 & 2 & 0' \\ 4 & 1 & 0^* & 5 \\ 5 & 5 & 0 & 6 \\ 4 & 0^* & 1 & 4 \end{pmatrix} & \begin{array}{c} + \\ \\ \\ \\ \end{array} \end{array} \quad 6.\begin{array}{cc} \begin{array}{cccc} \oplus & + & + & \end{array} & \\ \begin{pmatrix} 0^* & 8 & 6 & 0' \\ 0 & 1 & 0^* & 1 \\ 1 & 5 & 0 & 2 \\ 0 & 0^* & 1 & 0 \end{pmatrix} & \begin{array}{c} + \\ \\ \\ \\ \end{array} \end{array}$$

Помечаем его штрихом $c_{14}=0'$. Просматриваем строку $i=1$, в ней есть выделенный нуль $c_{11}=0^*$. Снимем выделение первого столбца знаком $\oplus$. Выделяем первую строку знаком (+). Получим матрицу (5). Если в этой матрице вычеркнуть выделенные столбцы $j=2, 3$ и выделенную строку $i=1$, то получим матрицу, в которой нет нулей, т. е. полученная матрица не содержит невыделенных нулей. Обозначим множество невыделенных строк через I_1. Очевидно, $I_1=\{2, 3, 4\}$, а множество невыделенных столбцов через J_1. Здесь

$J_1=\{1, 4\}$. Находим $min\{c_{ij}\}=4$, $i \in I_1$, $j \in J_1$. Прибавим число 4 к выделенным столбцам и вычтем его из невыделенных строк, получим матрицу 6. В результате такого преобразования получим новые невыделенные нули $c_{21} = 0$, $c_{41}=0$, $c_{44}=0$. Берем любой из них и отмечаем штрихом 0'. Например, $c_{21}=0'$. Так как в строке $i=2$ имеется 0^*, то снимаем выделение столбца $j=3$. Выделяем знаком (+) строку $i=2$. Перейдем к матрице 7. Среди невыделенных рядов матрицы есть невыделенный нуль. Это $c_{41}=0$. Штрихуем этот нуль $c_{41} = 0'$. В строке $i=4$ имеется 0^*. Снимаем выделение столбца $j=2$. Выделяем строку $i=4$. Получим матрицу 8. Среди невыделенных рядов находим невыделенный нуль $c_{33}=0$. Штрихуем этот нуль $c_{33}=0'$. В строке $i=3$ нет нулей со звездочкой. Поэтому нужно строить цепочку. От последнего нуля со штрихом $c_{33}=0'$ по столбцу переходим к нулю со звездочкой

$$7.\ \begin{array}{cc} \begin{matrix} \oplus & + & + & \end{matrix} & \\ \begin{pmatrix} 0^* & 8 & 6 & 0' \\ 0 & 1 & 0^* & 1 \\ 1 & 5 & 0 & 2 \\ 0 & 0^* & 1 & 0 \end{pmatrix} & \begin{matrix} + \\ + \\ \\ \end{matrix} \end{array} \qquad 8.\ \begin{array}{cc} \begin{matrix} \oplus & \oplus & \oplus & \end{matrix} & \\ \begin{pmatrix} 0^* & 8 & 6 & 0' \\ 0' & 1 & 0^* & 1 \\ 1 & 5 & 0' & 2 \\ 0' & 0^* & 1 & 0 \end{pmatrix} & \begin{matrix} + \\ + \\ \\ + \end{matrix} \end{array}$$

$c_{23}=0^*$. От $c_{23}=0^*$ по строке $i=2$ переходим к нулю со штрихом $c_{21}=0'$. От $c_{21}=0'$ по столбцу $j=1$ переходим к нулю со звездочкой $c_{11} = 0^*$. От $c_{11} = 0^*$ переходим по строке $i=1$ к $c_{14}=0'$. В столбце $j=4$ нет нулей со звездочкой. Следовательно, построение цепочки закончено. Построенная цепочка показана в матрице 9. На цепочке $c_{33}c_{23}c_{21}c_{11}c_{14}$ нули со штрихом заменяем на нули со звездочками. Звездочки над нулями, не находящимися на цепочке, сохраняем. Все остальные выделения уничтожаем. Переходим к матрице 10, в которой число независимых нулей увеличивается на единицу. На этом первая итерация закончена.

$$9.\ \begin{array}{cc} \begin{matrix} \oplus & \oplus & \oplus & \end{matrix} & \\ \begin{pmatrix} 0^* & 8 & 6 & 0' \\ 0' & 1 & 0^* & 1 \\ 1 & 5 & 0' & 2 \\ 0' & 0^* & 1 & 0 \end{pmatrix} & \begin{matrix} + \\ + \\ \\ + \end{matrix} \end{array} \qquad 10.\begin{pmatrix} 0 & 8 & 6 & 0^* \\ 0^* & 1 & 0 & 1 \\ 1 & 5 & 0^* & 2 \\ 0 & 0^* & 1 & 0 \end{pmatrix} \quad 11.\begin{pmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

Вторая итерация. Подсчитываем число независимых нулей: $k = 4$. Так как $n=4$, *то* $k=n$. Задача решена.

Оптимальный план назначения представлен в матрице 11.

7.5. ТРАНСПОРТНАЯ МОДЕЛЬ С ПРОМЕЖУТОЧНЫМИ ПУНКТАМИ (ДВУХЭТАПНАЯ ТРАНСПОРТНАЯ ЗАДАЧА)

Транспортная модель с промежуточными пунктами соответствует реальной ситуации, когда между исходными и конечными пунктами перевозок имеются промежуточные пункты для временного хранения грузов (транзит-ные пункты). Эта модель более общая, чем обычная транспортная, где пере-возки осуществляются непосредственно между пунктами отправления и назначения. В этом разделе показано, как транспортную модель с промежу-точными пунктами можно преобразовать в обычную транспортную задачу.

Допустим, имеется **m** *(i=1, m)* пунктов производства, *n (j=1, n)* пунктов потребления и *P (r=1, P)* – промежуточных баз. Как и в обычной транспортной задаче, обозначим через a_i, bj соответственно объемы поставок и потребления. Пусть d_r – мощность *r-й* базы, c_{ir} и c_{rj} – соответственно стоимость перевозки единицы продукции от поставщиков на базы и с баз к потребителям.

Тогда модель задачи примет вид $\min: Z=\sum_{i=1}^{m}\sum_{r=1}^{P}c_{ir}x_{ir}+\sum_{r=1}^{P}\sum_{j=1}^{n}c_{rj}x_{rj}$;

при ограничениях: $\sum_{r=1}^{P}x_{ir}\le a_i,\quad i=\overline{1,m}$;

Таблица 7.21

i \ j		
a_1 a_2 … a_i … a_m	$\lVert C_{ir}\rVert$	M
d_1 d_2 … d_r … d_P	0 0 M ⋱ 0 M ⋱ 0	$\lVert C_{rj}\rVert$

$\sum_{i=1}^{m}x_{ir}\le d_r,\quad r=\overline{1,P}$; $\sum_{r=1}^{P}x_{ir}\ge b_j,\quad j=\overline{1,n}$; $x_{ir}\ge 0;\ x_{rj}\ge 0.$

Если суммарная пропускная мощность баз равна суммарной мощности поставщиков и суммарному спросу потребителей, т. е. $\sum_{r=1}^{P} d_r = \sum_{i=1}^{m} a_i = \sum_{j=1}^{n} b_j$, то пропускные емкости баз будут использованы полностью и, следовательно, схема перевозок с баз к потребителям не зависит от схемы перевозок от поставщиков на базы. В таких условиях задачу можно решать по частям. Оптимальный план можно составить объединением плана поставок от поставщиков к базам и плана поставок с баз к потребителям. Однако оптимальный план двухэтапной транспортной задачи, вообще говоря, отличен от плана, полученного объединением оптимальных планов решения транспортной задачи для каждого этапа в отдельности.

Двухэтапную транспортную задачу легко свести к классической. Для этого базы будем считать одновременно поставщиками и потребителями. Для каждой базы в расширенной матрице (поставщики + базы) – (потребители + базы) отведем строку и столбец. Тогда матрица тарифов будет состоять из четырех блоков (табл. 7.21). В первом – левом верхнем блоке будем отражать связи поставщиков с базами *(ir)*, в четвертом – связи баз с потребителями. Второй – правый верхний блок показывает связи поставщиков с потребителями. Поскольку по условию задачи непосредственные перевозки от поставщиков к потребителям запрещены, то в этом блоке все тарифы считают равными *М* (где *М* – большое число). Третий – левый нижний блок образуется по строкам и столбцам базами, имеет форму квадрата. Так как перевозки между базами запрещаются, то соответствующие показатели с также считают равными *М*. В клетках третьего квадрата, в которых отражаются связи базы с самой собой, тарифы равны нулю. Поставки в этих клетках показывают величину неиспользованной мощности базы. Диагональ из нулевых тарифов, отражающая связи базы с самой собой, называется фиктивной.

Решение двухэтапной транспортной задачи имеет некоторые особенности. Основная из них – некоторое изменение нахождения базисного решения. Вначале необходимо распределить поставки в одном из блоков (первом или четвертом). Затем заполняется фиктивная диагональ и только потом распределяются поставки в другом блоке (четвертом или первом). Вторая особенность заключюючается в том, что если цикл пересчета проходит через фиктивную диагональ, то он обязательно проходит через нее дважды, одна вершина цикла, находящаяся на диагонали, будет всегда положительной, а другая – отрицательной.

7.6. ТРАНСПОРТНЫЕ СЕТИ

7.6.1. Задача о максимальном потоке. Обобщением транспортной задачи является понятие транспортной сети. В простой транспортной задаче разрешены перевозки только от поставщиков к потребителям. В реальных задачах некоторые пункты одновременно становятся поставщиками и потребителями. Примером такой задачи будет двухэтапная транспортная. Дальнейшее обобщение подобных задач ведет к понятию транспортной сети.

Прежде приведем некоторые определения из теории графов. Говорят, что задан граф (I, U), если задано непустое множество вершин I и множество дуг U. Причем каждый элемент u, $u \in U$ есть упорядоченная пара (ij) элементов множества I. Если оба множества I и U конечны, то говорят, что граф конечен. Геометрически множество I задается точками, а множество U – направленными отрезками, соединяющими эти точки. В материально-техническом снабжении и торговле вершинами графа можно считать поставщиков, потребителей или склады и базы, дуги – это связи между поставщиками, потребителями и базами.

Путем в графе назовем последовательность сцепленных дуг, позволяющих пройти из одной вершины в другую. Путь, у которого начальная вершина совпадает с конечной, называется контуром. Если элементам графа поставить в соответствие некоторые параметры, то получим сеть. Параметры могут быть заданы вершинам или дугам графа. Ориентированный граф превращается в сеть, если каждой вершине i поставлено в соответствие некоторое число d_i, называемое *интенсивностью* вершины, а каждой дуге – число r_{ij}, называемое *пропускной способностью* дуги. Функция r_{ij}, определенная на множестве дуг U, называется функцией пропускной способности, те вершины, для которых $d_i>0$, именуются источниками, вершины же, для которых $d_i<0$,– стоками. Если же $d_i = 0$, то вершина нейтральна. Например, для двухэтапной транспортной задачи источниками будут предприятия-поставщики, стоками – предприятия-потребители, нейтральные пункты – это промежуточные базы и склады, где отсутствует производство и потребление данного продукта. Для транспортной сети пропускные способности дуг – это максимальное количество груза, которое соответствующая коммуникация может пропустить за единицу времени. Отметим, что необязательно, чтобы пропускные способности ребер в разных направлениях были бы равны между собой, т. е. в общем случае $r_{ij} \neq r_{ij}$.

Рассмотрим сеть (I, U) с одним источником s и одним стоком s', на которой задана функция пропускной способности r_{ij}. Представим, что ребра (ij) этого графа – трубки с различными пропускными способностями. Если вершину s подключить к источнику, а вершину s' – к стоку, то по трубкам пойдет поток некоторой величины. Обозначим величину потока по каждому ребру через x_{ij}. В случае транспортной сети x_{ij} – это грузопоток из пункта i в пункт j. Таких чисел x_{ij}, очевидно, должно быть n^2, где n – число вершин графа. Задать произвольно n^2 чисел x_{ij} нельзя. Эти числа должны удовлетворять ряду ограничений. Прежде всего будем считать, что если поток из i в j равен x_{ij}, то поток из j в i равен x_{ji}, т. е.

$$x_{ij} = -x_{ji} \tag{7.6.1}$$

Если поток, прибывающий в пункт j, обозначается положительным числом, то поток, отправляющийся из пункта j, – отрицательным. Естественно считать, что $x_{11}=0$. Из физических соображений возникают следующие ограничения на числа x_{ij}.

1) Поток по любому ребру не может быть больше его пропускной способности:

$$0 \leq x_{ij} \leq r_{ij} \tag{7.6.2}$$

2) Для любой вершины, не являющейся ни источником, ни стоком, количество прибывающего вещества должно быть равно количеству отправляющегося. С учетом сделанного соглашения (1) последнее требование приводит к соотношению

$$\sum_i x_{ik} - \sum_j x_{kj} = 0, \quad \forall k \neq s, s'. \tag{7.6.3}$$

3) Так как вершина s – источник, а s' – сток, то общее количество вещества, вытекающего из источника s, совпадает с общим количеством вещества, притекающего в сток, т. е.

$$v = \sum_k x_{sk} = \sum_k x_{ks'} \tag{7.6.4}$$

Набор чисел x_i, удовлетворяющих условиям (7.6.1) – (7.6.4), называется *однородным потоком в сети*; величина V – *мощностью потока*. Поток $x_{ij}, \quad i, j = \overline{1, n}$ будем задавать матрицей $X = \left\| x_{ij} \right\|$. Очевидно, на главной диагонали этой матрицы стоят нули, а ниже главной диагонали – числа, имеющие противоположные знаки по отношению к числам, расположенным в матрице симметрично относительно главной диагонали. Возникает задача: определить величину максимального потока при заданной функции пропускной способности, т. е. найти x_{ij}, удовлетворяющие условиям (7.6.1) – (7.6.4), максимизирующие целевую функцию V, т. е.

$$\max : V = \sum_k x_{sk} = \sum_k x_{ks} \ .$$

Эта задача называется задачей о максимальном потоке.

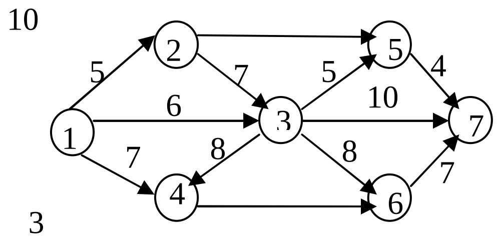


Рис. 7.2

7.6.2. Разрез сети. Критерий оптимальности потока. Разобьем все вершины сети с одним источником s и одним стоком s' на два непересекающиеся множества S и S' так, чтобы вершина s попала в множество S, а вершина s' – в множество S'. Такая разбивка множества именуется разрезом и обозначается (S, S').

В разрезе $s \in S, s' \in S', \ S \cap S' = \varnothing, S \cup S' = I$, где I – множество вершин сети. Величина $r\ (S,\ S') = \sum\limits_{i \in S} \sum\limits_{j \in S'} r_{ij}$ – пропускная способность разреза. Например, для графа, представленного на рис. 7.2, вершина *1* – источник, вершина 7– сток. Здесь можно построить несколько разрезов. Например, пусть $S_1 = \{1, 3\}$, тогда $S_1' = \{2, 4, 5, 6, 7\}$ и разрез определится множеством дуг: $(S_1,\ S_1') = =\{(1, 2), (1, 4), (3, 5), (3, 6), (3, 7)\}$. Если $S_2 = \{1, 2, 6\}$, тогда $S_2' = \{3, 4, 5, 7\}$ и, следовательно, $(S_2,\ S_2') = \{(1, 3), (1, 4), (2, 3), (2, 5), (6, 7)\}$. Найдем пропускные способности этих разрезов:

$$r\ (S_1,\ S_1') = r_{12} + r_{14+}\ r_{34+}\ r_{35+}\ r_{36} = 5 + 7 + 8 + 5 + 8 + 10 = 43.$$

$$r\ (S_2,\ S_2') = r_{13} + r_{14+}\ r_{23+}\ r_{25+}\ r_{67} = 6 + 7 + 7 + 10 + 7 = 37.$$

Отсюда видно, что различные разрезы могут иметь различную пропускную способность. Очевидно, должен существовать разрез с минимальной пропускной способностью. Назовем его минимальным. Из определения следует, что любой разрез блокирует все пути, идущие от источника s к стоку s'. Если мы рассмотрим в отдельности любой путь, то его пропускная способность не может превышать пропускной способности любой его дуги и он, следовательно, должен равняться пропускной способности дуги, имеющей минимальную пропускную способность.

Поскольку для любой дуги $x_{ij}, \leq r_{ij}$, то

$$\sum_{i \in S} \sum_{j \in S'} x_{ij} \leq \sum_{i \in S} \sum_{j \in S'} r_{ij} \tag{7.6.5}$$

т.е. величина потока через любой разрез не превосходит пропускной способности этого разреза.

Задача о максимальном потоке в сети формулируется так: при заданной топологии сети и известной пропускной способности дуг найти наибольшую величину потока и его распределение по дугам сети.

Для любого фиксированного потока можно указать несколько разрезов с различными пропускными способностями. Так как (7.6.5) справедливо для любого разреза, то увеличение потока возможно лишь до тех пор, пока он не достигает значения $\min r(S, S')$. Следовательно, величина потока x_{ij}, для которого

$$\sum_{i \in S} \sum_{j \in S'} x_{ij} = \min r(S, S'), \tag{7.6.6}$$

определяет максимальный поток сети. Отсюда *максимальный поток равен пропускной способности наименьшего разреза сети* (6).

РАЗДЕЛ 8. ЭЛЕМЕНТЫ ТЕОРИИ МАТРИЧНЫХ ИГР

8.1. ИГРЫ В ЧИСТЫХ СТРАТЕГИЯХ

Игра – это идеализированная математическая модель коллективного поведения: несколько участников влияют на ситуацию, причем их интересы различны. Столкновение противоположных интересов порождает конфликт, из которого требуется найти выход, наиболее устраивающий всех участников. Теория игр – это совокупность математических методов для анализа и оценки поведения в конфликтных ситуациях, а также выбора оптимальной стратегии поведения. Она широко используется в организации производства и экономике. Например, хозяйственная деятельность предприятий находится в тесной связи с поведением покупателя в отношении приобретения товаров. Каждое предприятие принимает конкретное решение относительно своей деятельности, учитывая, как будут вести себя покупатели, как изменится их численность, активность.

Под игрой будем понимать последовательность действий (ходов), которая осуществляется по некоторым правилам. Конфликтующие стороны условно называются *игроками*. Цель игры – разработка таких предложений игрокам, чтобы тот, кто выигрывает, сумел выиграть как можно больше, а тот, кто проигрывает, проиграл бы как можно меньше. *Ходом* в игре назовем выбор одного из возможных действий и его осуществление. *Стратегией* игрока назовем систему правил, однозначно определяющих выбор поведения игрока на каждом ходе в зависимости от ситуации, сложившейся в процессе игры. Каждая фиксированная стратегия, которую может выбрать игрок, называется его *чистой стратегией*. Теория игры дает указания игрокам о выборе стратегий, которые обеспечат максимально возможный средний выигрыш.

Для составления математической модели игры примем следующие допущения:

1. рассматриваем только *парные* игры, с участием двух игроков A и B, причем для определенности будем считать, что игрок A всегда выигрывает, а B проигрывает;
2. рассматриваем игры с *нулевой суммой*, т.е. в сумме выигрыш A и проигрыш B равны нулю, или, что тоже самое, выигрыш A равен проигрышу B;

3. все ходы игроков и результат игры можно выразить количественно, в виде платы за ход;
4. стратегии каждого игрока известны заранее;
5. игра состоит из двух ходов: игрок A выбирает одну из своих стратегий A_i $(i = 1,m)$, а игрок B – Bj $(j = 1,n)$, причем каждый выбор осуществляется при полном незнании выбора другого игрока.

Пусть A и B – два игрока, располагающие конечным числом возможных действий – чистых стратегий (выраженными количественно), соответственно, $A_1,A_2,...,A_m$ и $B_1,B_2,...,B_n$. Игрок A может выбрать любую чистую стратегию A_i $(i = \overline{1,m})$, в ответ на которую игрок B может выбрать любую свою чистую стратегию B_j $(j = \overline{1,n})$. Выбор пары стратегий (A_i,B_j) однозначно определяет результат a_{ij} – выигрыш игрока A и проигрыш игрока B. Очевидно, если $a_{ij}<0$, игрок A проигрывает. При известных значениях a_{ij} выигрыша для каждой пары (A_i,B_j) чистых стратегий, можно составить матрицу выигрышей игрока A (проигрышей игрока B), которую называют *платежной матрицей*, или *матрицей игры:*

$$\begin{array}{c} \\ A_1 \\ A_2 \\ \cdots \\ A_m \end{array} \begin{array}{c} \begin{matrix} B_1 & B_2 & \dots & B_n \end{matrix} \\ \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \end{array}$$

в которой строки соответствуют чистым стратегиям игрока A, а столбцы – чистым стратегиям игрока B.

Во время игры игрок A, выбирая чистую стратегию A_i, ничего не знает о том, как будет действовать игрок B, поэтому выбирая строку, он определяет *гарантированный выигрыш (т.е. минимальный)* по каждой стратегии и выбирает в каждой строке минимальный элемент. После этого он выбирает ту строку, в которой этот минимальный элемент наибольший. Такой выбор при любых действиях игрока B обеспечит игроку A максимальный гарантированный выигрыш. Таким образом, выбор оптимальной стратегии игрока A определяется числом $a = \max_i \left(\min_j a_{ij} \right)$,которое называется чистой *нижней ценой игры* (максимином), а соответствующая стратегия максиминной.

Игрок B выбирает стратегию (соответствующий столбец B_j) так, чтобы проиграть как можно меньше. Он не знает как будет действовать его партнер, поэтому в каждом столбце ищет максимальный элемент, то есть то, что ему придется заплатить при самом неблагоприятном для него выборе игрока A. Желая минимизировать свой проигрыш, он выбирает тот столбец, где будет минимум *гарантированного (т.е. максимального) проигрыша*. Таким образом, выбор оптимальной стратегии игрока B определяется числом: $\beta = \min\limits_j \max\limits_i a_{ij}$, которое называется чистой *верхней ценой игры* (минимаксом), а соответствующая стратегия минимаксной. Очевидно, что всегда $\alpha \leq \beta$. Если $a = \beta = \nu$, то ν называют ценой игры, элемент a_{ij}, в таком случае называют *седловой точкой*.

Если платежная матрица имеет седловую точку, то чистые стратегии уравновешены и ни одному из игроков нет смысла уклоняться от своей оптимальной стратегии. Решением игры будет (A_i,B_j,ν). Ситуации равновесия определяют по схеме:

$$\left.\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}\begin{matrix} \to \min\limits_j a_{1j} \\ \to \min\limits_j a_{2j} \\ \to \min\limits_j a_{ij} \\ \to \min\limits_j a_{mj} \end{matrix}\right\} \max_i \min_j a_{ij} = \alpha$$

$$\underbrace{\begin{matrix} \downarrow & \downarrow & \downarrow \\ \max\limits_i a_{i1} & \max\limits_i a_{i2} & \max\limits_i a_{in} \end{matrix}}_{\min\limits_j \max\limits_i a_{ij} = \beta}$$

Пример 1.

$$\left.\begin{pmatrix} 1 & 3 & 5 & 7 \\ 2 & 3 & 6 & 8 \\ (5) & 9 & 8 & 10 \\ 4 & 6 & 7 & 8 \end{pmatrix}\begin{matrix} \to 1 \\ \to 2 \\ \to 5 \\ \to 4 \end{matrix}\right\} \alpha = 5$$

$$\underbrace{\begin{matrix} \downarrow & \downarrow & \downarrow & \downarrow \\ 5 & 9 & 8 & 10 \end{matrix}}_{\beta = 5}$$

Цена игры $\nu = a = \beta = 5$. Седловая точка $a_{31} = 5$. Решение $(A_3, B_1, 5)$.

Пример 2. Магазин может завезти в различных пропорциях товары трех типов (A_1, A_2, A_3). Их реализация, а следовательно, и получаемая прибыль (a_{ij}) зависят от вида товара и состояния спроса. Предполагая, что последний может характеризоваться тремя состояниями (B_1, B_2, B_3) и учитывая, что спрос связан с изменением моды и прогнозирование его невозможно, определить оптимальную стратегию в закупке товаров из условия средней гарантированной прибыли при следующей матрице прибылей

$$
\begin{array}{cc}
 & \begin{array}{ccc} B_1 & B_2 & B_3 \end{array} \\
\begin{array}{c} A_1 \\ A_2 \\ A_3 \end{array} & \left.\begin{bmatrix} 20 & 15 & 10 \\ 16 & 12 & 13 \\ 13 & 18 & (13) \end{bmatrix} \begin{array}{l} \to 10 \\ \to 12 \\ \to 13 \end{array} \right\} \to \alpha = \max(\min) a_{ij} = 13
\end{array}
$$

$$\downarrow \quad \downarrow \quad \downarrow \qquad \max \qquad \alpha = \beta = \nu = 13$$

$$\min \underbrace{20 \quad 18 \quad 13}_{\substack{\downarrow \\ \beta = \min(\max)\, a_{ij} = 13}}$$

Оптимальной стратегией для магазина в данных условиях будет закупка товаров вида A_3, которая при любом состоянии спроса обеспечит гарантированную прибыль 13 ден.ед. Решение игры: $(A_3, B_3, 13)$.

8.2. МАТРИЧНЫЕ ИГРЫ В СМЕШАННЫХ СТРАТЕГИЯХ

Если точки равновесия в платежной матрице нет, игра ведется в смешанных стратегиях. Смешанную стратегию образует вероятная смесь разных стратегий, она содержит в себе все чистые стратегии, взятые с соответствующим уровнем вероятности.

Смешанной стратегией игрока A называется вектор $\vec{X} = (x_l, x_2, \ldots, x_m)$, компоненты которого являются *вероятностями* выбора игроком соответствующей стратегии и, очевидно, удовлетворяют условиям

$$x_i \geq 0 \ (i = \overline{1, m}); \quad \sum_{i=1}^{m} x_i = 1.$$

Смешанной стратегией игрока B называют аналогичный вектор

$$\vec{Y} = (y_l, y_2, \ldots, y_n), \text{ где } y_j \geq 0 \ (j = \overline{1, n}); \quad \sum_{j=1}^{n} y_j = 1.$$

При использовании смешанных стратегий игра приобретает случайный характер и величина выигрыша игрока *A* (проигрыша игрока *B)* также является случайной величиной, функцией смешанных стратегий $\vec{X}$ и $\vec{Y}$ и определяется по формуле $f(\vec{X},\vec{Y})=\sum_{i=1}^{m}\sum_{j=1}^{n}a_{ij}x_i y_j$,которую называют *платежной функцией или функцией выигрыша*.

Смешанные стратегии $\vec{X}*$ и $\vec{Y}*$ называют оптимальными, если они образуют седловую точку для платежной функции $f(\vec{X},\vec{Y})$, то есть удовлетворяют неравенству $f(\vec{X},\vec{Y}*)\le f(\vec{X}*,\vec{Y}*)\le f(\vec{X}*,\vec{Y})$

это равносильно $f(\vec{X}*,\vec{Y}*)=\min_y \max_x f(\vec{X},\vec{Y})=\max_x \min_y f(\vec{X},\vec{Y})$

Величину $f(\vec{X}*,\vec{Y}*)$ называют ценой игры.

Поиск оптимальных смешанных стратегий начинают с упрощения платежной матрицы. Если в платежной матрице элементы *k-ой* строки не меньше соответствующих элементов *s*-ой строки, т. е. $a_{kj}\ge a_{sj}$ $(j=\overline{1,n})$, то говорят, что стратегия A_k доминирует над стратегией A_s. Аналогично, если элементы *l-го* столбца не превосходят соответствующих элементов *r*-го столбца, то есть $a_{il}<a_{ir}$ $(i=\overline{1,m})$, то говорят, что стратегия B_l доминирует над стратегией B_r. Частным случаем доминирования является дублирование стратегий, когда $a_{ki}=a_{sj}$ $(j=\overline{1,n})$ или $a_{il}=a_{ir}$ $(i=\overline{1,m})$. Исключение из платежной матрицы доминирующих стратегий (ими игроку пользоваться заведомо невыгодно) позволяет уменьшить ее размер, а это упрощает решение игры. Вероятность применения доминирующих стратегий равна нулю.

Оптимальные смешанные стратегии $\vec{X}*$ и $\vec{Y}*$в игре с платежной матрицей $A=\lfloor a_{ij}\rfloor_{m\times n}$ и ценой ν остаются оптимальными и для игры с платежной матрицей $\underset{b>0}{\left[ba_{ij}+c\right]_{m\times n}}$, и ценой ν. На этом основании платежную матрицу можно всегда преобразовать так, что ее элементы будут целыми неотрицательными числами.

Пример. Упростить платежную матрицу $\begin{pmatrix} 1 & 2 & -3 & -1 & 0 \\ 3 & 4 & -1 & 2 & 1 \\ 2 & 0 & 1 & -3 & 6 \\ 3 & 4 & -1 & 2 & 1 \\ 1 & -2 & 1 & -4 & 3 \end{pmatrix}$

Элементы 2-ой и 4-ой строк равны; элементы 1-ой строки меньше соответствующих элементов 2-ой, а элементы 5-ой строки не превосходят соответствующих элементов 5-ей. Это значит, что игроку *A*, который хочет максимизировать свой выигрыш, лучше применить стратегии A_2 и A_3, а не стратегии A_1 и A_5. Отбросим доминирующие стратегии и получим матрицу

$$\begin{pmatrix} 3 & 4 & -1 & 2 & 1 \\ 2 & 0 & 1 & -3 & 6 \end{pmatrix}.$$

Она имеет элементы 1-го и 2-го столбцов больше соответствующих элементов 4-го, а элементы 5-го столбца больше соответствующих элементов 3-го. Поэтому столбцы B_1,B_2 и B_3 можно отбросить. (Игроку *B*, стремящемуся минимизировать свой проигрыш, выгоднее иметь дело со стратегиями B_3 и B_4). Получим матрицу $\begin{pmatrix} -1 & 2 \\ 1 & -3 \end{pmatrix}$, которая уже не подлежит дальнейшему упрощению.

8.3. ГЕОМЕТРИЧЕСКОЕ РЕШЕНИЕ МАТРИЧНЫХ ИГР

Игры с платежными матрицами размеров $2\times n$ и $m\times 2$ можно решать геометрически. Если нет седловой точки, то решение такой игры представляется смешанными оптимальными стратегиями, а цена игры ν принадлежит интервалу $(\alpha;\beta)$. Рассмотрим такое решение на примерах.

Пример 1. Найти решение игры, заданной матрицей $\begin{pmatrix} 2 & 5 \\ 6 & 4 \end{pmatrix}$ и дать геометрическую интерпретацию.

Прежде всего, проверим наличие седловой точки. Определим: $\alpha=max(2,4)=4$; $\beta=min(6,5)=5$ и $4\le v\le 5$. Седловой точки нет, поэтому игра идет в смешанных стратегиях.

Использование оптимальной стратегии $\vec{X}$ $=(x,,x_2)$ обеспечивает игроку *A* получение среднего выигрыша ν при любых стратегиях игрока *B*, то есть

$$\begin{cases} a_{11}x_1 + a_{21}x_2 = v & \left(\text{\textit{при стратегии} } B_1\right); \\ a_{12}x_1 + a_{22}x_2 = v & \left(\text{\textit{при стратегии} } B_2\right); \\ x_1 + x_2 = 1 \end{cases} \quad \text{или} \quad \begin{cases} 2x_1 + 6x_2 = v, \\ 5x_1 + 4x_2 = v, \\ x_1 + x_2 = 1. \end{cases}$$

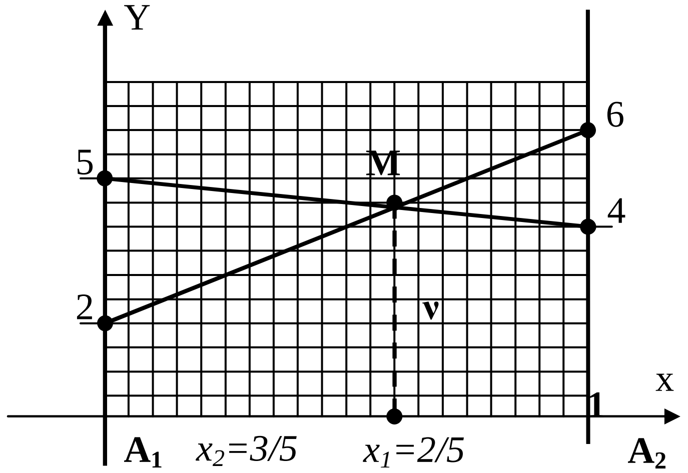


Рис.8.1

Аналогично, для стратегии $\vec{Y} = (y_1, y_2)$ игрока B при любых стратегиях игрока A:

$$\begin{cases} \mathrm{a}_{11}y_1 + a_{12}y_2 = v, \\ \mathrm{a}_{21}y_1 + a_{21}y_2 = v, \\ y_1 + y_2 = 1. \end{cases} \quad \text{или} \quad \begin{cases} 2y_1 + 5y_2 = v, \\ 6y_1 + 4y_2 = v, \\ y_1 + y_2 = 1. \end{cases}$$

Решая полученные системы уравнений, получим оптимальные смешанные стратегии $\vec{X} = (2/5;\ 3/5)$, $\vec{Y} = (1/5; 4/5)$ и цену игры $v = 22/5$.

Геометрическая интерпретация (Рис.8.1). Построим прямоугольную систему координат на плоскости, и на оси OX отложим отрезок A_1A_2, равный единице. Левый конец отрезка соответствует стратегии A_1, правый – стратегии A_2. Промежуточные точки отрезка с координатами $0<x<1$ соответствуют некоторым смешанным стратегиям $\vec{X} = (x_1, x_2)$ где $x_1 = 1-x$; $x_2 = x$. На концах отрезка проведем прямые, параллельные оси OY, на которых будем откладывать выигрыш игрока A при его соответствующих чистых стратегиях. Так, при применении игроком B стратегии B_1 игрок A выиграет 2 при применении A_1 или 6 при применении A_2. Соединим эти точки прямой линией. Аналогично: при стратегии B_2 откладываем точки 5 (A_1) и 4 (A_2) и соединяем прямой. Каждая точка полученных прямых имеет ординату, равную цене игры при выборе игроком A смешанной стратегии $(1-x,\ x)$. Но игрок A руководствуется максиминной стратегией, т.е. выбирает из минимальных выигрышей максимальный. Минимальные выигрыши определяются ординатами точек, принадлежащих нижней ломаной линии (жирно выделена) – *нижняя граница игры.* Из всех точек нижней границы игры выбираем т. *M* с наибольшей ординатой, которая и будет ценой игры.

Координаты т. *M* получим, решая систему

$$\begin{cases} 2x_1 + 6x_2 = \nu, \\ 5x_1 + 4x_2 = \nu, \Rightarrow \mathrm{x}_1 = 2/5;\ x_2 = 3/5;\ \nu = 22/5. \\ x_1 + x_2 = 1. \end{cases}$$

Аналогично, найдем оптимальную стратегию игрока *B,* для этого составим систему:

$$\begin{cases} 2y_1 + 5y_2 = 22/5, \\ 6y_1 + 4y_2 = 22/5, \\ y_1 + y_2 = 1. \end{cases}$$

Для решения которой можно выбрать любые два уравнения. y_1 = 1/5; y_2=4/5. Ответ: Оптимальная смешанная стратегия игрока *A:* $\vec{X}$ = *(2/5; 3/5*), игрока *B:* $\vec{Y}$ *=(1/5;4/5)*, цена игры ν *= 22/*5*=4,4.* Результат можно проверить по рисунку.

Пример 2. Найти решение игры, заданной матрицей $A = \begin{pmatrix} 1 & 4 & 3 \\ 4 & -1 & 0 \end{pmatrix}$.

Проверим матрицу на седловую точку*:* $\alpha=max(1,-1)=1;\ \beta=min(4,4,3)=3,\ 1\le\nu\le3.$ Седловой точки нет. Игра ведется в смешанных стратегиях. Так как матрица имеет размер *2×4,* поиск оптимальных смешанных стратегий начинаем с игрока *A* – у него две чистые стратегии. Строим прямые *1-4, 4-(-1), 3-0* (Рис. 8.2).

Жирно выделена линия, определяющая нижнюю границу игры. Точка с наибольшей ординатой на этой линии – *M.* Она лежит на пересечении прямых *1-4* и *3-0,* которые соответствуют стратегиям B_1 и B_3 игрока *B*. Можно сразу сделать вывод, что *активными* у игрока *B* будут стратегии B_1 и B_3, т.е. вероятность их применения будет отлична от нуля, а вероятность применения стратегии B_2 равна нулю ($y_1\neq0;\ y_2=0;\ y_3\neq0,$ т.е. $y_1+y_3=1$). Этот вывод используем при решении задачи для игрока *B*.

Для игрока *A* составим систему уравнений, отражающую тот факт, что он получит свой средний выигрыш при использовании игроком *B* как первой, так и третьей стратегии:

$$\begin{cases} x_1 + 4x_2 = \nu, \\ 3x_1 \quad\quad = \nu, \\ x_1 + x_2 = 1. \end{cases}$$

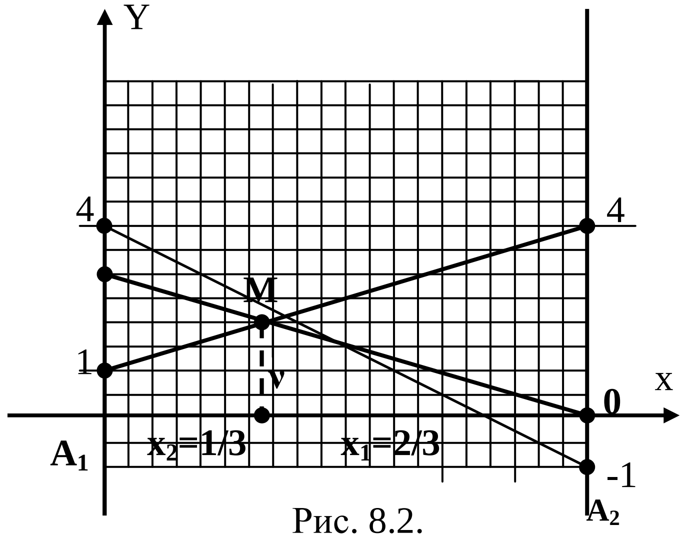


Рис. 8.2.

Решим эту систему: $\begin{cases} x_1 + 4x_2 = 3x_1, \\ x_1 + x_2 = 1. \end{cases}$ $\begin{cases} 2x_2 = x_1, \\ 3x_2 = 1. \end{cases}$ $x_2=1/3;\ x_1=2/3;\ \nu=2.$

Для игрока *B* система уравнений:

$$\begin{cases} y_1 + 3y_3 = \nu, \\ 4y_1 \quad\quad = \nu, \\ y_1 + y_3 = 1. \end{cases} \text{ или } \begin{cases} y_1 + 3y_3 = 2, \\ 4y_1 \quad\quad = 2, \\ y_1 + y_3 = 1. \end{cases} \quad y_1=1/2;\ y_3=1/2;\ \nu=2.$$

На рисунке можно проверить этот результат. Итак, для игрока *A*: $\vec{X} = \left(2/3;\ 1/3\right)$; для игрока *B:* $\vec{Y} = \left(1/2;\ 0;\ 1/2\right)$; цена игры $\nu=2$.

Пример 3. Решить матричную игру, заданную матрицей

$$A = \begin{pmatrix} 4 & 3 \\ 2 & 4 \\ 0 & 5 \\ -1 & 6 \end{pmatrix} \begin{matrix} \to 3 \\ \to 2 \\ \to 0 \\ \to -1 \end{matrix} \to \max = 3$$

$$\begin{matrix} \downarrow & \downarrow \\ 4 & 6 \end{matrix}$$

Найдем $\alpha=max(3;2;0;-1)=3$ *и* $\beta=min(4;6)=4 \Rightarrow 3<\nu<4$. Седловой точки нет. Так как матрица имеет размер 4×2, решение ищем сначала для игрока *B*. Сделаем рисунок 8.3, чтобы определить *верхнюю границу игры*, так как игрок *B* выбирает из максимальных проигрышей минимальный (макси-минная стратегия).

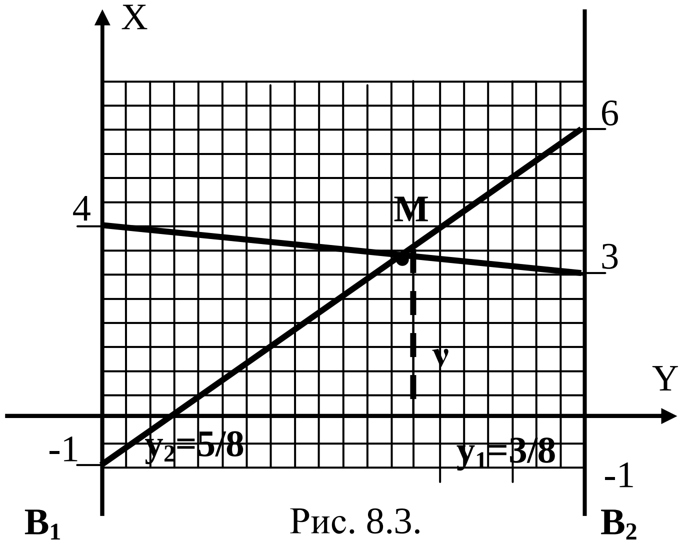


Рис. 8.3.

Жирной линией выделена верхняя граница игры, M – точка с *min* ординатой, которая соответствует цене игры. Активными стратегиями игрока *A* будут *1-я* и *4-я*. Это означает, что вероятности выбора *2-й* и *3-й* стратегий равны нулю, т.е. $x_2=0; x_3=0$.

Решение для игрока *B* получим из следующей системы:

$$\begin{cases} 4y_1 + 3y_2 = \nu, \\ -y_1 + 6y_2 = \nu, \Rightarrow y_1 = 3/8;\ y_2 = 5/8;\ \nu = 27/8. \Rightarrow \vec{Y} = (3/8;\ 5/8). \\ y_1 + y_2 = 1. \end{cases}$$

Вероятность выбора *1-й* стратегии игроком *B*: $y_1=3/8$; второй – $y_2 = 5/8$, цена игры $\nu = 27/8 = 3\ ^3/_8$. Теперь найдем x_1 и x_4 из любых двух уравнений из трех возможных:

$$\begin{cases} 4x_1 - x_4 = 27/8, \\ 3x_1 + 6x_4 = 27/8,\ x_1 = 7/8;\ x_4 = 1/8; \quad \vec{X} = (7/8;\ 0;\ 0;\ 1/8). \\ x_1 + x_2 = 1. \end{cases}$$

Итак, для игрока *A*: $\vec{X} = (7/8;\ 0;\ 0;\ 1/8)$, для игрока *B*: $\vec{Y} = (3/8;\ 5/8)$, цена игры $\nu = 3\,3/8$.

8.4. СВЕДЕНИЕ МАТРИЧНОЙ ИГРЫ К ЗАДАЧЕ ЛИНЕЙНОГО ПРОГРАММИРОВАНИЯ

Пусть игра задана платежной матрицей: $\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$

Оптимальные смешанные стратегии $\vec{X}^* = (x_1^*, x_2^*, ..., x_n^*)$ и $\vec{Y}^* = (y_1^*, y_2^*, ..., y_n^*)$ игроков *A* и *B* могут быть найдены в результате решения пары двойственных задач линейного программирования.

прямая задача

$$\min F = \sum_{i=1}^{m} x_i$$

$$\begin{cases} a_{11}x_1 + a_{21}x_2 + \cdots + a_{m1}x_m \geq 1, \\ a_{12}x_1 + a_{22}x_2 + \cdots + a_{m2}x_m \geq 1, \\ \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots, \\ a_{1n}x_1 + a_{2n}x_2 + \cdots + a_{mn}x_m \geq 1, \\ x_i \geq 0, \quad i = \overline{1, m} \end{cases}$$

двойственная задача

$$\max Z = \sum_{j=1}^{n} y_j$$

$$\begin{cases} a_{11}y_1 + a_{12}y_2 + \cdots + a_{1n}y_n \leq 1, \\ a_{21}y_1 + a_{22}y_2 + \cdots + a_{2n}y_n \leq 1, \\ \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots, \\ a_{m1}y_1 + a_{m2}y_2 + \cdots + a_{mn}y_n \leq 1, \\ y_j \geq 0, \quad j = \overline{1, n} \end{cases}$$

Так как эти задачи образуют двойственную пару задач линейного программирования, нет необходимости решать обе. Получив решение одной из них, достаточно воспользоваться соответствием между переменными двойственных задач, и из строки оценок целевой функции последней симплекс-таблицы выписать значения компонент оптимального вектора другой задачи.

После чего найдем решение матричной игры

$$\nu = \frac{1}{F_{\min}} = \frac{1}{Z_{\max}}; \quad x_i^* = \nu \cdot x_i, (i = \overline{1, m}), \quad y_j^* = \nu \cdot y_j, (j = \overline{1, n}).$$

Пример 1. Предприятие выпускает скоропортящуюся продукцию, которую оно может сразу отправить потребителю (стратегия A_1), отправить на склад для хранения (стратегия A_2) или подвергнуть дополнительной обработке (стратегия A_3) для длительного хранения. В свою очередь потребитель может немедленно приобрести эту продукцию (стратегия B_1), приобрести ее в течение небольшого отрезка времени (стратегия B_2) или затребовать ее после длительного периода времени (стратегия B_3). Если предприятие выберет стратегию A_1, то дополнительные затраты на хранение и обработку не потребуются. Однако, если при этом потребитель применит стратегию B_2 или тем более B_3, то предприятие потерпит убытки из-за порчи части продукции. Наоборот, если предприятие выберет стратегию A_2, а потребитель стратегию B_1, то возникнут неоправданные расходы на консервацию продукции. Определить оптимальную стратегию предприятия при следующей матрице затрат.

$$\begin{array}{c} \\ A_1 \\ A_2 \\ A_3 \end{array} \begin{array}{c} \begin{array}{ccc} B_1 & B_2 & B_3 \end{array} \\ \begin{pmatrix} 2 & 5 & 8 \\ 7 & 6 & 10 \\ 12 & 10 & 8 \end{pmatrix} \end{array}$$

Проверим наличие седловой точки $\alpha=8;\ \beta=12;\ a\neq\beta$. Седловой точки нет, $8\leq v\leq 12$. Решим задачу методом линейного программирования. Составим пару двойственных задач.

$F=x_1+x_2+x_3\rightarrow min$	$z=y_1+y_2+y_3\rightarrow max$
$\begin{cases} 2x_1+7x_2+12x_3\geq 1, \\ 5x_1+6x_2+10x_3\geq 1, \\ 8x_1+10x_2+8x_3\geq 1 \\ x_i\geq 0, \quad i=\overline{1,3} \end{cases}$	$\begin{cases} 2y_1+5y_2+8y_3\leq 1, \\ 7y_1+6y_2+10y_3\leq 1, \\ 12y_1+10y_2+8y_3\leq 1 \\ y_i\geq 0, \quad i=\overline{1,3} \end{cases}$

Удобнее решать задачу на максимум. Приведем ее к каноническому виду и составим симплекс-таблицу.

$$z=y_1+y_2+y_3\rightarrow max$$

$$\begin{cases} 2y_1+5y_2+8y_3+y_4 = 1, \\ 7y_1+6y_2+10y_3+y_5 = 1, \\ 12y_1+10y_2+8y_3+y_6 = 1, \\ y_i\geq 0, \quad i=\overline{1,6} \end{cases}$$

Базис	$C_Б$	$\vec{A}$	y_1	y_2	y_3	y_4	y_5	y_6	ϑ
y_4	0	1	2	5	8	1	0	0	1/8
y_5	0	1	7	6	$\boxed{10}$	0	1	0	1/10-min
y_6	0	1	12	10	8	0	0	1	1/8
Δ_j		0	-1	-1	$\underline{-1}$	0	0	0	
y_4	0	2/10	-36/10	2/10	0	1	-8/10	0	1
y_3	1	1/10	7/10	6/10	1	0	1/10	0	1/6
y_6	0	2/10	64/10	$\boxed{52/10}$	0	0	-8/10	1	1/26-min
Δ_j		1/10	-3/10	$\underline{-4/10}$	0	0	1/10	0	
y_4	0	11/65	-436/130	0	0	1	-10/13	1/26	
y_3	1	1/26	32/26	0	1	0	5/26	-3/26	
y_2	1	1/26	32/26	1	0	0	-4/26	5/26	
Δ_j		3/26	5/26	0	0	0	1/26	2/26	

Все $\Delta_j \geq 0$ => получено оптимальное решение одной из пары двойственных задач: $Z_{max} = 3/26;$ $\vec{Y} = \left(0;\ {}^{1}\!/_{26};\ {}^{2}\!/_{26};\ {}^{11}\!/_{65};\ 0;\ 0\right)$ или $\vec{Y} = \left(0;\ {}^{1}\!/_{26};\ {}^{2}\!/_{26}\right)$. Устанавливая соответствие между переменными и Δ_j, получаем для второй задачи $f_{\min} = {}^{3}\!/_{26};\ \vec{X} = \left(0;\ {}^{1}\!/_{26};\ {}^{2}\!/_{26}\right)$. Теперь можно получить решение матричной игры.

Цена игры $\nu = {}^{1}\!/_{f_{\min}} = {}^{1}\!/_{Z_{\max}} = {}^{26}\!/_{3};$

$$\vec{X}^* = \nu \cdot \vec{X} = \left(0;\ {}^{1}\!/_{3};\ {}^{2}\!/_{3}\right);\quad \vec{Y}^* = \nu \cdot \vec{Y} = \left(0;\ {}^{1}\!/_{3};\ {}^{2}\!/_{3}\right)$$

Предприятию не выгодно использовать стратегию A_1, стратегию A_2 оно выберет с вероятностью *1/3*, а стратегию A_3 – с вероятностью *2/3*.

Пример 2. Бригада рабочих, которая к следующей весне должна построить электростанцию, живет в поселке. В связи с надвигающейся зимой возникла проблема угольных запасов для отопления поселка. Если зима будет мягкой, потребуется *12 тыс. т* угля, если нормальной – *15 тыс. т,* а если суровой *–18 тыс. т.* Весной поселок переезжает на новый объект строительства и излишки угля будут потеряны. В зависимости от типа зимы стоимость тонны угля составляет соответственно *10; 12; 14 ден.ед.* В настоящее время можно приобрести уголь по *10 ден.ед.* за тонну, а остальное, если потребуется, позже. Найти оптимальный выход из сложившейся ситуации.

Бригада может выбрать одну из трех возможных стратегий:

A_1 – закупить *12 тыс. т* угля, тогда в случае мягкой зимы расход составит $12\times10=120$ *тыс.ден.ед.;* в случае нормальной – $12\times10+3\times12=156;$ суровой – $12\times10+\ +6\times14=204$ *тыс. ден.ед.;*

A_2 – закупить *15 тыс. т* угля, тогда в случае мягкой или нормальной зимы расход составит $15\times10=150$ *тыс.ден.ед.;* суровой – $15\times10+3\times14=192$ *тыс.ден.ед.;*

A_3 – закупить *18 тыс. т* угля, тогда в любом случае расход составит $18\times10=180$ *тыс.ден.ед.*

Зима может быть: B_1 – мягкой; B_2 – нормальной; B_3 – суровой.

В платежной матрице знак минус у всех элементов, т.к. в любом случае бригада несет расходы, их надо минимизировать. Существует седловая точка, соответствующая запасу в *18 тыс. т.* угля и расходу *180 тыс.ден.ед.,* т.е. дополнительные закупки невыгодны.

$$\begin{pmatrix} -120 & -156 & -204 \\ -150 & -150 & -192 \\ -180 & -180 & -180 \end{pmatrix} \begin{matrix} \to -204 \\ \to -192 \\ \to -180 \end{matrix}$$

$$\begin{matrix} \downarrow & \downarrow & \downarrow \\ -120 & -150 & -180 \end{matrix}$$

Пример 3. Швейная фабрика по плану должна в апреле израсходовать *35 т.ден.ед.* на пошив брюк и костюмов, причем брюки обходятся ей по *10*, а костюмы по *25 ден.ед.* Реализация продукции будет происходить в мае по ценам: брюки *20*, костюм *45 ден.ед.* По статистическим данным, в случае прохладной погоды можно продать *500* брюк и *1200* костюмов, а в случае теплой погоды *2000* брюк и *600* костюмов. Товар, не реализованный в течение месяца, дохода не приносит (учитывая расходы на содержание, переоценку и т.д.). Как максимизировать средний доход фабрики? Для фабрики имеется две стратегии: A_1 – шить одежду в расчете на теплую погоду. Тогда при прохладной погоде прибыль фабрики $500\times(20\text{-}10)\text{-}1500\times(20\text{-}10)+600\times(45\text{-}25)=2$ *м .ден. ед.*, а при продаже при теплой погоде $2000\times(20\text{-}10)+600\times(45\text{-}25)=32$ *м. ден .ед.* A_2 – шить одежду в расчете на прохладную погоду. Тогда при прохладной погоде прибыль фабрики $500\times(20\text{-}10)+1200\times(45\text{-}25)=29$ *м. ден. ед.*, а при продаже при теплой погоде $500\times(20\text{-}10)+600\times(45\text{-}25)\text{-}\ 600\times25=2$ *м. т. ден.ед.*

У погоды в мае, месяце продажи, также две стратегии: B_1 – прохладная, B_2 – теплая.

Матрица игры $\begin{pmatrix} 2 & 32 \\ 29 & 2 \end{pmatrix}$ не имеет седловой точки, $\alpha=2;\ \beta=29;\ 2\le\nu\le29.$

Геометрическая иллюстрация решения не обязательна, т.к. у обеих игр 0-стратегии. Решим задачу для игрока *А*, т.е. швейной фабрики.

$$\begin{cases} 2x_1 + 29x_2 = \nu, \\ 32x_1 + 2x_2 = \nu, \\ x_1 + x_2 = 1. \end{cases} \Rightarrow \begin{matrix} x_1 = 9/19, \\ x_2 = 10/19, \\ \nu = 308/19 \cong 16{,}2. \end{matrix}$$

Средняя прибыль фабрики составит *16,2 тыс.ден.ед.* при применении ею стратегий в отношении *9:10*.

РАЗДЕЛ 9. НЕЛИНЕЙНАЯ ОПТИМИЗАЦИЯ

9.1. ПОСТАНОВКА ЗАДАЧ НЕЛИНЕЙНОЙ ОПТИМИЗАЦИИ

В нелинейном программировании рассматривается оптимизация моделей задач, в которых либо ограничения, либо целевая функция, либо то и другое нелинейны.

Классификация основных методов нелинейной оптимизации:

1. Аналитические, использующие классические методы дифференциального и вариационного исчисления. Они применяются при отсутствии ограничений и при их наличии типа равенств и (или) неравенств.
2. Численные.
3. Графические, базирующиеся на графическом представлении функций, подлежащих максимизации или минимизации.
4. Методы исследования возможных вариантов, основанные на идее генерирования возможных вариантов с целью выбора наилучшего из них.
5. Экспериментальные, они в настоящее время выделены в новое направление – математическую теорию планирования эксперимента.

Рассмотрим основные понятия, связанные с экстремумом нелинейной функции цели. В зависимости от наличия или отсутствия ограничений, накладываемых на переменные $x_j,\ j=\overline{1,n}$, а также от особенностей целевой функции $Z=z(X)$ будем различать следующие виды экстремумов (максимумов или минимумов): безусловный абсолютный экстремум; безусловный относительный экстремум; условный абсолютный экстремум; условный относительный экстремум. Для простоты совокупность неизвестных $x_j,(\ j=\overline{1,n})$ обозначим через X и будем считать X точкой в n-мерном пространстве.

Если в задаче математического программирования отсутствуют ограничения, то говорят, что $z(X)$ в точке X^o имеет безусловный абсолютный или глобальный экстремум: максимум, если $z(X^o)\geq z(X), \forall X$. В случае $z(X^o)\leq z(X)$ говорят об абсолютном (глобальном) минимуме. Если неравенство $z(X^o)\geq z(X)$ выполняется не для всех X, а только для точек, лежащих в некоторой ε- окрестности точки X^o ($\varepsilon>0$), то говорят об относительном (локальном) максимуме. Аналогично вводится понятие относительного (локального) минимума.

Рассмотрим задачу математического программирования в общей постановке:

$$\max_{X \in \Omega}(\min): Z = z(X),$$

где Ω – область допустимых управлений. Целевая функция Z=z(X) имеет условный абсолютный максимум в точке $X^o \in \Omega$, если для любых $X \in \Omega$ справедливо утверждение $z(X^o) \geq z(X)$. Если же неравенство $z(X^o) > z(X)$ справедливо в некоторой ε-окрестности точки X^o – это условный относительный максимум. Аналогично вводится понятие условного абсолютного и относительного минимума. В случае, если указанные в определении экстремумов неравенства выполняются как строгие неравенства, – речь идет о сильном экстремуме, если же неравенства нестрогие, – о слабом.

Особенности задач нелинейного программирования: для выяснения трудностей, порождаемых нелинейностью, сопоставим задачи линейного и нелинейного программирования. Отметим характерные особенности линейных задач, которые позволяют строить универсальные алгоритмы их решения и которые отсутствуют в нелинейных задачах.

Область допустимых планов – это выпуклое множество с конечным числом угловых (крайних) точек. Напомним, что крайней называется всякая точка множества, которая не является внутренней ни для какого отрезка, целиком принадлежащего этому множеству. По крайней мере в одной из угловых (крайних) точек целевая функция достигает экстремального значения. Такую точку можно найти за конечное число шагов. В задачах же нелинейного программирования множество допустимых планов может быть невыпуклым, несвязным, иметь бесконечное число крайних точек.

Экстремального значения линейная целевая функция достигает в одной из крайних точек (на границе области допустимых решений). При нелинейной целевой функции экстремум может достигаться не только на границе, но и внутри области допустимых решений.

Экстремальное локальное значение целевой функции является в то же время экстремальным глобальным значением. В задачах же нелинейного программирования целевая функция в области допустимых решений может иметь несколько локальных экстремумов.

9.2. ЗАДАЧИ БЕЗУСЛОВНОЙ ОПТИМИЗАЦИИ

9.2.1. Классический метод нахождения экстремумов функций

Чтобы найти точки экстремума функции, необходимо решить систему уравнений

$$\frac{\partial f(X)}{\partial x_j} = 0, \quad j = \overline{1, n}. \tag{9.2.1}$$

Эти условия лежат в основе всех классических методов нахождения экстремумов функций. Точка, для которой они выполняются, называется стационарной. Это необходимые условия существования max(min) непрерывной и гладкой функции. В случае функции одной переменной легко получить достаточные условия экстремума. Напомним их. Если при переходе через стационарную точку X^o производная $f'(X)$ меняет положительный знак на отрицательный, то в точке X^o функция $f(X)$ имеет максимум. Если же $f'(X)$ меняет отрицательный знак на положительный, то в точке X^o функция $f(X)$ имеет минимум. В случае, когда смены знака производной $f'(X)$ не происходит, в точке X^o нет ни максимума, ни минимума. При нахождении экстремумов функции одной переменной широко применяется также второй признак существования экстремума. Если X^o есть стационарная точка функции $f(X)$ и $f''(X^o)<0$, то в точке X^o функция $f(X)$ имеет максимум, а если $f''(X^o)>0$, то функция $f(X)$ в точке X^o имеет минимум. Итак, обращение в нуль частных производных первого порядка – необходимое условие существования экстремума.

На первый взгляд кажется, что использование этого метода дает возможность просто определить оптимум нелинейной функции нескольких переменных. Однако это не так. Можно указать ряд трудностей. Прежде всего, при большом числе переменных возникают трудности при решении системы уравнений (9.2.1); Кроме того, необходимые условия существования экстремума не достаточны. Системе уравнений могут удовлетворять точки перегиба, седловые точки и др. Учет достаточных условий нахождения экстремумов функции многих переменных сложный как в алгебраическом, так и в вычислительном плане. Так, для функции одной переменной достаточ-ными условиями существования минимума (максимума) становится положительный (отрицательный) знак производной четного порядка:

$$\frac{\partial^{2k} f(X)}{\partial x^{2k}} > 0, \left(\frac{\partial^{2k} f(X)}{\partial x^{2k}} < 0 \right)$$

В случае функции двух переменных достаточным условием существования экстремума будет положительная определенность матрицы Δ:

$$\Delta = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_2^2} \end{pmatrix} > 0$$

Если Δ >0, то в исследуемой стационарной точке X^o функция $f\ (x_1,\ x_2)$ имеет экстремум, а именно: максимум при $\frac{\partial^2 f(X^o)}{\partial x_1^2} < 0$, и минимум – при $\frac{\partial^2 f(X^o)}{\partial x_1^2} > 0$. Если же Δ <0, то в стационарной точке экстремума нет. Когда Δ=0, то для решения вопроса приходится привлекать производные высшего порядка. Аналогичные достаточные условия можно сформулировать для функции *n* переменных. Кроме того, классический метод дифференциального исчисления дает возможность найти экстремум только в том случае, если он лежит внутри области определения функции. Если экстремум находится на границе области определения – этот метод становится бессильным. Классический метод предполагает отсутствие дополнительных уравнений или неравенств, связывающих между собой часть переменных, и др.

9.2.2. Сведение задачи с ограничениями к задаче безусловной оптимизации.

Пусть требуется отыскать план *X*, доставляющий экстремальное значение целевой функции, т. е.

$$\max(\min): Z=Z(X) \tag{9.2.2}$$

и удовлетворяющий системе ограничений в форме равенств

$$f_i(X) = b_i, \quad i = \overline{1,m} \tag{9.2.3}$$

Для решения этой задачи применим метод множителей Лагранжа, который играет фундаментальную роль в науке оптимизации решений, применим к задачам с ограничениями-равенствами при отсутствии требования неотрицательности и целочисленности переменных. Он допускает обобщения задачи нелинейного программирования с ограничениями в виде равенств.

Рассмотрим простейший вариант задачи, когда число переменных равно двум и число ограничений − одному:

$$\max (\min): Z = z(x_1, x_2); \tag{9.2.4}$$

$$f(x_1, x_2) = b. \tag{9.2.5}$$

Пусть функции $Z(x_1, x_2)$ и $f(x_1, x_2)$ непрерывны и, по крайней мере, дважды дифференцируемые. Решим уравнение (9.2.5) относительно перемен-ной x_2, получим выражение

$$x_2 = \varphi(x_1). \tag{9.2.6}$$

Подставим (9.2.6) в целевую функцию (9.2.4), тогда

$$Z = z(x_1, \varphi(x_1)). \tag{9.2.7}$$

Найти экстремальное значение функции (9.2.7), в которую уже включено исходное ограничение (9.2.5), можно из уравнения $Z = z(x, \varphi)$ – неявная функция, поэтому можно записать так:

$$\frac{dz\left(x_1, \varphi\left(x_1\right)\right)}{dx_1} = \frac{\partial Z}{\partial x_1} + \frac{\partial Z}{\partial \varphi} + \frac{\partial \varphi}{\partial x_1} \tag{9.2.8}$$

Продифференцируем также функцию $f(x_1, x_2) = b$ как неявную:

$$f'_{x_1}(x_1, x_2) + f'_{x_2}(x_1, x_2) \cdot \frac{dx_2}{dx_1} = 0\,.$$

откуда, учитывая(9.2.6), имеем:

$$\frac{d\varphi}{dx_1} = -\frac{f'_{x_1}(x_1, x_2)}{f'_{x_2}(x_1, x_2)} = -\frac{\partial f / \partial x_1}{\partial f / \partial x_2} \tag{9.2.9}$$

Из (9.2.7), (9.2.8) и (9.2.9) получим уравнение для определения x_1^o:

$$\frac{dZ}{dx_1} = \frac{\partial Z}{\partial x_1} - \frac{\partial Z}{\partial x_2}\frac{\partial f / \partial x_1}{\partial f / \partial x_2} = 0. \tag{9.2.10}$$

Обозначим

$$\frac{\partial Z / \partial x_2}{\partial f / \partial x_2} \equiv \lambda. \tag{9.2.11}$$

Тогда (9.2.10) примет вид:

$$\frac{\partial Z}{\partial x_1} - \lambda \frac{\partial f}{\partial x_1} = 0\,. \tag{9.2.12}$$

Кроме того, из (9.2.11) имеем:

$$\frac{\partial Z}{\partial x_2} - \lambda \frac{\partial f}{\partial x_2} = 0\,. \tag{9.2.13}$$

Таким образом, необходимые условия существования условного локального экстремума можно представить в виде системы трех уравнений:

$$\begin{cases} \dfrac{\partial Z}{\partial x_1} - \lambda \dfrac{\partial f}{\partial x_1} = 0; \\ \dfrac{\partial Z}{\partial x_2} - \lambda \dfrac{\partial f}{\partial x_2} = 0; \\ f(x_1, x_2) = b. \end{cases} \qquad (9.2.14)$$

Решив систему уравнений (9.2.14) с неизвестными $X = (x_1, x_2)$ и λ, найдем все точки X^o, подозрительные на экстремум.

9.2.3. Метод неопределенных множителей Лагранжа

Основное удобство представления необходимых условий в виде системы (9.2.14) состоит в том, что эти уравнения можно получить чисто формальным путем. Для этого составим вспомогательную функцию:

$$L(X, \lambda) = z(x_1,x_2) + \lambda[b-f(x_1,x_2)],$$

которая называется функцией Лагранжа, а множитель λ – неопределенным множителем Лагранжа. Если в ней считать x_1, x_2 и λ независимыми переменными, найти частные производные по x_1, x_2 и λ и приравнять их к нулю, то будем иметь необходимые условия (9.2.14) для определения точек, подозрительных на экстремум.

Пусть классическая задача об условном экстремуме сформулирована в общем виде (9.2.2) – (9.2.3). Построим функцию Лагранжа:

$$L(x_1, x_2, ..., x_n, \lambda_1, \lambda_2, ..., \lambda_m) = z(x_1, x_2, ..., x_n) + \sum_{i=1}^{m} \lambda_i [b_i - f_i(x_1, x_2, ..., x_n)].$$

Обозначим через X и λ совокупности переменных x_j $j = \overline{1,n}$ и λ_i, $i = \overline{1,m}$,тогда

$$L(X, \Lambda) = z(X) + \sum_{i=1}^{m} \lambda_i [b_i - f_i(X)] \qquad (9.2.15)$$

Исходная задача отыскания условного экстремума задачи (9.2.4) – (9.2.5) заменяется задачей отыскания безусловного экстремума функции Лагранжа $L(X, \Lambda)$, которая может быть решена путем приравнивания к нулю частных производных по x_j $j = \overline{1,n}$ и λ_i, $i = \overline{1,m}$:

$$\frac{\partial L(X, \Lambda)}{\partial x_j} = \frac{\partial\, z(X)}{\partial x_j} - \sum_{i=1}^{m} \lambda_i \frac{\partial f_i(X)}{\partial x_j} = 0, \ j = \overline{1,n}; \qquad (9.2.16)$$

$$\frac{\partial L(X,\Lambda)}{\partial \lambda_i} = b_i - f_i(X) = 0, i = \overline{1,m}. \qquad (9.2.17)$$

Условия (9.2.16)– (9.2.17) необходимы для экстремума задачи (9.2.2) – (9.2.3). В самом деле, пусть X^o– экстремальный план задачи. Тогда $f_i(X^o)=b_i$ $i=\overline{1,m}$. Поэтому из (9.2.15) следует: $L(X^o, \Lambda)=Z(X^o)$, значит,

$$\frac{\partial L(X^o,\Lambda)}{\partial x_j} \equiv \frac{\partial z(X^o)}{\partial x_j} = 0,\ j = \overline{1,n},$$

т. е. каждая точка X^o, в которой $z(X)$ достигает экстремального значения, должна быть решением системы (9.2.16) – (9.2.17). Однако условия (9.2.16)–(9.2.17) не являются достаточными.

Пример. Найти оптимальное распределение ограниченного ресурса в a единиц между n потребителями, если прибыль, получаемая при выделении j-му потребителю x_j единиц ресурса, вычисляется по формуле $c_j\sqrt{x_j}$. Составим математическую модель задачи

$$\max : Z = \sum_{j=1}^{n} c_j \sqrt{x_j}$$

при ограничении $\quad \sum_{j=1}^{n} c_j x_j = a$

Формируем функцию Лагранжа: $\quad L(X,\lambda) = \sum_{j=1}^{n} c_j\sqrt{x_j} + \lambda(a - \sum_{j=1}^{n} x_j)$

Приравниваем ее частные производные по x_j, $j=\overline{1,n}$ и λ к нулю:

$$\frac{\partial L(X,\Lambda)}{\partial x_j} = \frac{c_j}{2\sqrt{x_j}} - \lambda = 0,$$

$$\frac{\partial L(X,\Lambda)}{\partial \lambda} = a - \sum_{j=1}^{n} x_j = 0$$

Решив эту систему уравнений, получим:

$$x_j^* = \frac{c_j^2 a}{\sum_{j=1}^{n} c_j^2}, \quad j = \overline{1,n}; \quad \lambda^* = \sqrt{\frac{\sum_{j=1}^{n} c_j^2}{4a}}.$$

Интерпретация множителей Лагранжа. Будем считать, что правые части системы ограничений (9.2.3) могут в некоторых пределах изменяться тогда точка, доставляющая целевой функции экстремальное значение, тоже будет изменяться. Она станет функцией вектора b: $x_j(b) = x_j(b_1, b_2, ..., b_m), \quad j = \overline{1,n}$.

Рассмотрим теперь функцию Лагранжа как функцию вектора ресурсов b:

$$L(b) = z\left(X(b)\right) + \sum_{i=1}^{m} \lambda_i(b)[b_i - f_i\left(X(b)\right)] \tag{9.2.18}$$

Продифференцируем функцию (9.2.18) по b_i,

$$\frac{\partial L(b)}{\partial b_i} = \left(\frac{\partial Z}{\partial x_j} - \sum_{i=1}^{m} \lambda_i \frac{\partial f_i}{\partial x_j}\right)\frac{\partial x_j}{\partial b_i} + \sum_{i=1}^{m}[b_i - f_i\left(X(b)\right)]\frac{\partial \lambda_i}{\partial x_i} + \lambda_i, \quad i = \overline{1,m}. \tag{9.2.19}$$

Учитывая условия(9.2.16) – (9.2.17) из (9.2.19) получим:

$$\frac{\partial L(b)}{\partial b_i} = \lambda_i, \quad i = \overline{1,m}.$$

Для оптимального плана $L(X^o, \Lambda) \equiv z(X^o)$, следовательно, $\frac{\partial z(X)}{\partial b_i} = \lambda_i$.

Отсюда вывод: множители Лагранжа показывают, как реагирует целевая функция на изменение соответствующих параметров b_i. По величине X^o можно выяснить, какой вклад в достижение цели вносит изменение каждого параметра b_i: $\frac{\partial z(X)}{\partial b_i} \approx \frac{\Delta Z}{\Delta b_i} \approx \lambda_i$. Отсюда при $\Delta bi = 1$ $\Delta Z \approx \lambda_i$.

9.2.4. Выпуклые множества и выпуклые функции

Под *прямой* $П(X_1\ X_2)$, проходящей через точки X_1 и X_2, будем по-нимать геометрическое множество точек, удовлетворяющих уравнению $Z=(1-\lambda)\cdot X_1+$ $+\lambda X_2$, где λ – любое действительное число. Под отрезком $O(X_1, X_2)$ с концами в точках X_1 и X_2 понимают множество $Z=(1-\lambda)X_1+\lambda X_2$, где $0 \le \lambda \le 1$. Множество Ω называется выпуклым, если оно вместе с любыми своими двумя точками X_1 и X_2 содержит соединяющий их отрезок $O(X_1, X_2)$. Легко показать, что если множество выпукло, то оно вместе с любыми своими точками $X_1, X_2, ..., X_к$ содержит и их выпуклую линейную комбинацию:

$$X = \sum_{i=1}^{k} \lambda_i X_i, \quad \lambda_i \ge 0, \quad \sum_{i=1}^{k} \lambda_i = 1.$$

Функция $f(X)$, заданная на множестве Ω, называется выпуклой вниз, если множество Ω выпукло и для любых $X_1 \in \Omega$, $X_2 \in \Omega$ и $0<\lambda<1$ выполняется неравенство

$$f((1-\lambda)\,X_1+\lambda\,X_2) \le (1-\lambda)f(X_1)+\lambda f(X_2).$$

Функция же $f(X)$, заданная на множестве Ω, называется выпуклой вверх, если множество Ω выпукло и при любых $X_1 \in \Omega$, $X_2 \in \Omega$ и $0<\lambda<1$ выполняется неравенство

$$f((1-\lambda) X_1+\lambda X_2)\geq(1-\lambda)f(X_1)+\lambda f(X_2).$$

Методом математической индукции, исходя из определений выпуклых функций, можно доказать более общие утверждения (неравенство Иенсена): если $f(X)$ – выпуклая вниз функция, определенная на выпуклом множестве Ω, то при любых $X_1, X_2, ..., \in \Omega\ X_k$ выполняется неравенство

$$f\left(\sum_{i=1}^{k}\lambda_i X_i\right)\leq\sum_{i=1}^{k}\lambda_i f(X_i),\quad \lambda_i\geq 0,\quad \sum_{i=1}^{k}\lambda_i=1.$$

Если $f(X)$ – выпуклая вверх функция на этом же множестве Ω, то при любых $X_1, X_2, ..., X_k \in \Omega$ справедливо неравенство

$$f\left(\sum_{i=1}^{k}\lambda_i X_i\right)\geq\sum_{i=1}^{k}\lambda_i f(X_i),\quad \lambda_i\geq 0,\quad \sum_{i=1}^{k}\lambda_i=1.$$

Справедливы следующие утверждения:

1) Если заданная на выпуклом множестве Ω функция $f(X)$ выпукла вниз, то при любом действительном X множество точек, удовлетворяющих неравенству $f(X)\leq b$, будет выпуклым.
2) Если заданная на выпуклом множестве Ω функция $f(X)$ выпукла вверх, то при любом действительном b будет выпуклым множество точек удовлетворяющих неравенству $f(X)\geq b$.

   Пересечением двух множеств называется множество, состоящее из общей части данных множеств. Очевидно, пересечение двух выпуклых множеств Ω_1 и Ω_2 есть выпуклое множество.

   Из утверждений (1) – (2) следует, что если на выпуклом множестве Ω заданы выпуклые вниз функции $f_i(X)$, $i=\overline{1,m}$, то при любых действительных b_i, $i=\overline{1,m}$, множество точек, удовлетворяющих системе неравенств $f_i(X)\leq bi$, $i=\overline{1,m}$, будет выпуклым множеством. Если функции $f_i(X)$, $i=\overline{1,m}$ выпуклые вверх, то множество точек, удовлетворяющих системе ограничений $f_i(X)\geq b_i$, $i=\overline{1,m}$, будет выпуклым множеством.
3) Если определенные на одном и том же выпуклом множестве Ω функции $f_i(X)$, $i=\overline{1,m}$ выпуклы вниз (вверх), то их линейная комбинация $f(X)=\sum_{i=1}^{m}\lambda_i f_i(X)$, $\lambda_i>0, i=\overline{1,m}$ будет также функцией, выпуклой вниз (вверх).

4) Чтобы функция $f(X)$ на выпуклом множестве была выпуклой вверх (вниз), необходимо и достаточно, чтобы такой она была по каждой своей переменной при произвольных фиксированных значениях остальных переменных.

9.2.5. Задачи выпуклого программирования. Теорема Куна – Таккера

В нелинейном программировании особое место принадлежит задачам выпуклого программирования.

Определение. Выпуклым программированием называется раздел нелинейного программирования, предметом которого является изучение методов отыскания экстремума выпуклой вниз (вверх) функции на выпуклом множестве.

Экстремумы выпуклых функций. Глобальным максимумом (минимумом) функции $f(X)$ на множестве $X \in \Omega$ называется такая точка $X^o \in \Omega$, что для любой точки $X_к \in \Omega$ $f(X^o) \geq f(X)$ (соответственно $f(X^o) \leq f(X)$). Если это неравенство выполняется в некоторой ε-окрестности точки X^o, то такая точка доставляет функции локальный максимум (минимум).

Справедливы следующие **теоремы**.

1) Каждая точка локального максимума выпуклой вверх функции на выпуклом множестве Ω является точкой глобального максимума.
2) Каждая точка локального минимума выпуклой вниз функции $f(X)$ на выпуклом множестве Ω является точкой глобального минимума.

Экстремум выпуклой вверх (вниз) функции может быть и не единственным.

Отметим некоторые свойства точек максимума выпуклой вверх функции.

6. Множество точек максимума выпуклой вверх функции, заданной на выпуклом множестве, выпукло.
7. Если $f(X)$ – строго выпуклая вверх функция, заданная на выпуклом множестве, то ее максимум единственный.

С помощью приведенных выше определений трудно установить выпуклость или вогнутость функции. В связи с этим используются и другие критерии, в, частности, следующий: дважды дифференцируемая функция $F(x)$ является строго вогнутой в окрестности точки X^o, если все нечетные определители вида

$$F_{11}(X^o), \qquad \begin{vmatrix} F_{11}(X^o) & F_{12}(X^o) \\ F_{21}(X^o) & F_{22}(X^o) \end{vmatrix},$$

$$\begin{vmatrix} F_{11}(X^o) & F_{12}(X^o) & F_{13}(X^o) \\ F_{21}(X^o) & F_{22}(X^o) & F_{23}(X^o) \\ F_{31}(X^o) & F_{32}(X^o) & F_{33}(X^o) \end{vmatrix} \quad \text{и т.д.}$$

отрицательны, при этом определители четного порядка являются положительными; и строго выпуклой, если все указанные выше определители (и четного и нечетного порядка) положительны. Здесь $F_{ij}(X^o)$– частные производные второго порядка в точке X^o.

Пример. Исследовать функцию $Z = x_1^2 - x_1x_2 + 5x_2^2$ на выпуклость (вогнутость) в окрестности точки X^o. Составим и вычислим определители

$$F_{11}(X^o) = 2 > 0, \qquad \begin{vmatrix} F_{11}(X^o) & F_{12}(X^o) \\ F_{21}(X^o) & F_{22}(X^o) \end{vmatrix} = \begin{vmatrix} 2 & -1 \\ -1 & 10 \end{vmatrix} = 22 > 0.$$

Замечание. Сумма выпуклых (вогнутых) на некотором множестве функций является выпуклой (вогнутой) функцией.

Пусть дана задача нелинейного программирования с ограничениями-неравенствами:

$$\max : Z = z(x);$$

$$f_i(X) \le b_i, \quad i = \overline{1,m};$$

$$x_j \ge 0, \quad j = \overline{1,n}$$

Подобно задаче с ограничениями-равенствами без условия неотрицательности запишем для данной задачи функцию Лагранжа:

$$L(X, \Lambda) = z(x) + \sum_{i=1}^{m} \lambda_i [b_i - f_i(X)], \qquad (9.2.20)$$

где λ_i, $i = \overline{1,m}$ – неопределенные множители Лагранжа. Функция $L(X,\Lambda)$ имеет в точке (X^0, Y^0) седловую точку, если выполняются соотношения $L(X,\Lambda^0) \le L(X^0,\Lambda^0) \le L(X^0,\Lambda)$ для всех X и Λ из ε-окрестности точки (X^0, Y^0) $(\varepsilon > 0)$. Справедлива **теорема**. Чтобы функция двух векторных переменных (9.2.20) имела седловую точку, необходимо и достаточно выполнения следующих *условий Куна–Таккера:*

$$\frac{\partial}{\partial x_j} L(X^0, \Lambda^0) \leq 0;$$

$$x_j^0 \frac{\partial}{\partial x_j} L(X^0, \Lambda^0) = 0, \quad x_j^0 \geq 0;$$

$$\frac{\partial}{\partial \lambda_i} L(X^0, \Lambda^0) \geq 0;$$

$$\lambda_i^0 \frac{\partial}{\partial x_i} L(X^0, \Lambda^0) = 0, \quad \lambda_i^0 \geq 0 .$$

Если решается задача минимизации, то функция Лагранжа будет иметь седловую точку, если выполняются соотношения:

$L(X^0,\Lambda) \leq L(X^0,\Lambda^0) \leq L(X,\Lambda^0)$. Условия же Куна–Таккера существования седловой точки функции Лагранжа перепишутся в виде:

$$\frac{\partial}{\partial x_j} L(X^0, \Lambda^0) \geq 0;$$

$$x_j^0 \frac{\partial}{\partial x_j} L(X^0, \Lambda^0) = 0, \quad x_j^0 \geq 0;$$

$$\frac{\partial}{\partial \lambda_i} L(X^0, \Lambda^0) \leq 0;$$

$$\lambda_i^0 \frac{\partial}{\partial x_i} L(X^0, \Lambda^0) = 0, \quad \lambda_i^0 \geq 0 .$$

Аналог условий Куна-Таккера иногда формулируется и доказывается для недифференцируемых функций $Z(X)$ и $f_i(X)$. В этом случае теорему о существовании седловой точки функции Лагранжа называют теоремой Куна- Таккера. Чтобы сформулировать ее для задачи максимизации

$$\max : Z = z(x); \tag{9.2.21}$$

$$f_i(X) \leq b_i, \quad i = \overline{1, m}; \tag{9.2.22}$$

где $Z(X)$ и $f_i(X)$ – выпуклые вверх на выпуклом множестве Ω функции, введем понятие *условий Слейтера*. Говорят, что ограничения (9.2.22) удовлетворяют условию Слейтера, если существует вектор X^* Є Ω, такой, что

$$f_i(X^*) \leq b_i, \quad i = \overline{1, m}; \tag{9.2.23}$$

Теорема Куна – Таккера. Пусть ограничения (9.2.22) задачи (9.2.21) – (9.2.22) удовлетворяют условию Слейтера(9.2.23). Чтобы вектор X^0 Є Ω был решением задачи (9.2.21) – (9.2.22) необходимо и достаточно, чтобы существовал вектор $\Lambda^0 \geq 0$, такой, что пара (X^0, Λ^0) составляла седловую точку функции Лагранжа, т. е. для всех $X\in\Omega$ и $\Lambda\geq 0$ должны выполняться условия $L(X,\Lambda^0)\leq L(X^0,\Lambda^0)\leq L(X^0,\Lambda)$.

9.2.6. Квадратичное программирование

Пример. применения условий Куна-Таккера – квадратичное программмирование. Это раздел нелинейного программирования, когда целевая функция квадратична, а ограничения – линейные равенства и неравенства.

Для задач квадратичного программирования характерны следующие особенности.

1) Целевая функция $Z(X)$ в общем случае представляет собой сумму линейной и квадратичной форм:

$$z(X)=\sum_{j=1}^{n}c_jx_j+\sum_{k=1}^{n}\sum_{j=1}^{n}d_{kj}x_kx_j.$$

2) Функции системы ограничений $f_i(x_1,x_2,...,x_n),\quad i=\overline{1,m}$ линейны относительно переменных $x_j,\quad j=\overline{1,n}$.

3) Ограничения могут иметь вид равенств и неравенств и включать требование неотрицательности всех или части переменных x_j.

4) Коэффициенты c_j и a_{ij} в выражениях Z и f_j, а также свободные члены b_i могут быть произвольными действительными числами, чего нельзя сказать о коэффициентах d_{kj}, на которые накладываются определенные условия. В задачах отыскания минимума квадратичная форма должна быть выпукла вниз, т.е. положительно полуопределенной; на максимум – выпукла вверх, т. е. отрицательно полуопределенной. Таким образом, задача квадратичного программирования в общем виде запишется так:

$$\max(\min):Z=\sum_{j=1}^{n}c_jx_j+\sum_{k=1}^{n}\sum_{j=1}^{n}d_{kj}x_kx_j \qquad (9.2.24)$$

при ограничениях

$$\sum_{j=1}^{n}a_{ij}x_j\leq b_i,\quad i=\overline{1,u} \qquad (9.2.25)$$

$$\sum_{j=1}^{n}a_{ij}x_j\geq b_i,\quad i=\overline{u+1,v} \qquad (9.2.26)$$

$$\sum_{j=1}^{n}a_{ij}x_j=b_i,\quad i=\overline{v+1,m} \qquad (9.2.27)$$

$$x_j\geq 0,\quad j=\overline{1,n} \qquad (9.2.28)$$

где max относится к отрицательной, а min – к положительной полуопределенности квадратичной формы. Так как целевая функция либо выпукла вверх (задача на max), либо вниз (задача на min) и множество допустимых решений (9.2.25) – (9.2.28) выпукло, то локальный экстремум задачи квадратичного программирования будет его глобальным экстремумом.

Вводя дополнительные переменные (как и в задаче линейного программирования), систему ограничений (9.2.25) – (9.2.26) можно свести к ограничениям-равенствам. Тогда задача квадратичного программирования примет вид:

$$\max(\min): Z = \sum_{j=1}^{n} c_j x_j + \sum_{k=1}^{n}\sum_{j=1}^{n} d_{kj} x_k x_j \tag{9.2.29}$$

$$\sum_{j=1}^{n} a_{ij} x_j = b_i, \quad i = \overline{1, m} \tag{9.2.30}$$

$$x_j \geq 0, \quad j = \overline{1, n} \tag{9.2.31}$$

или в матричной форме:

$$\max(\min): Z = C^T X + X^T DX;$$
$$AX = b, \quad X \geq 0.$$

Для получения необходимых и достаточных условий оптимальности плана задачи квадратичного программирования рассмотрим задачу максимизации. Тогда квадратичная форма $X^T DX$ выпукла вверх, т.е. отрицательно определена.

Используем условия Куна – Таккера для получения необходимых и достаточных условий оптимальности плана задачи (9.2.29) – (9.2.31). Составим для этого функцию Лагранжа:

$$L(X, \Lambda) = \sum_{j=1}^{n} c_j x_j + \sum_{k=1}^{n}\sum_{j=1}^{n} d_{kj} x_k x_j + \sum_{i=1}^{m} \lambda_i \left(b_i - \sum_{j=1}^{n} a_{ij} x_j \right). \tag{9.2.32}$$

Условия Куна – Таккера примут вид:

$$\frac{\partial}{\partial x_j} L(X^0, \Lambda^0) = c_j + 2\sum_{k=1}^{n} d_{kj} x_k^0 - \sum_{i=1}^{m} \lambda_i^0 a_{ij} \leq 0; \tag{9.2.33}$$

$$x_j^0 \left(c_j + 2\sum_{k=1}^{n} d_{kj} x_k^0 - \sum_{i=1}^{m} \lambda_i^0 a_{ij} \right) = 0; \tag{9.2.34}$$

$$x_j^0 \geq 0; \tag{9.2.35}$$

$$\frac{\partial}{\partial \lambda_i} L(X^0, \Lambda^0) = b_i - \sum_{j=1}^{n} a_{ij} x_i^0 \geq 0; \tag{9.2.36}$$

$$\lambda_i^0\left(b_i-\sum_{j=1}^{n}a_{ij}x_i^0\right)=0\,; \tag{9.2.37}$$

$$\lambda_i^0\geq 0; \tag{9.2.38}$$

Таким образом, если X^0 – оптимальный план задачи квадратичного программирования, то должен существовать такой вектор Λ^0, что выполняются условия(9.2.33) –(9.2.38). Сведем систему условий (9.2.33) к каноническому виду, т. е. введением вектора дополнительных переменных

$$v^*=\sum_{i=1}^{m}\lambda_i^0a_{ij}-2\sum_{k=1}^{n}d_{kj}x_k^0-c_j$$

запишем их в виде равенств. Тогда условия оптимальности плана задачи квадратичного программирования запишутся так:

$$c_j+2\sum_{k=1}^{n}d_{kj}x_k^*-\sum_{i=1}^{m}\lambda_i^*a_{ij}+v_j^*=0\,;$$

$$x_j^*v_j^*=0,\quad x_j^*\geq 0,\quad v_j^*\geq 0.$$

Так как условия–ограничения представлены в виде равенств(9.2.30), то на вектор Λ условие неотрицательности не накладывается; условия же (9.2.37) –(9.2.38) будут автоматически выполняться при любом допустимом решении, и их можно исключить.

Итак, имеем окончательно: любое решение *X≥0, V≥0,* удовлетворяющее системе уравнений

$$\sum_{j=1}^{n}a_{ij}x_j=b_i; \tag{9.2.39}$$

$$2\sum_{k=1}^{n}d_{kj}x_k^*-\sum_{i=1}^{m}\lambda_i^*a_{ij}+v_j^*=-c_j; \tag{9.2.40}$$

$$x_j^*v_j^*=0, \tag{9.2.41}$$

$$x_j^*\geq 0,\quad v_j^*\geq 0. \tag{9.2.42}$$

дает вектор $x_j\geq 0,\quad j=\overline{1,n}$, являющийся оптимальным решением задачи квадратичного программирования с отрицательно полуопределенной квадратичной формой. В матричной форме система уравнений (9.2.39) –(9.2.41) примет вид:

$$\begin{gathered}AX=b\,;\\ 2DX-A^T\Lambda+V=-C^T;\\ X^TV=0\end{gathered} \tag{9.2.43}$$

Неотрицательные решения $X \geq 0$, $V \geq 0$ системы уравнений (9.2.43) дают решение задачи квадратичного программирования. Если эта система имеет единственное решение, то и задача квадратичного программирования имеет единственное решение, если бесчисленное множество или несовместна, то тоже можно сказать и о решении исходной задачи.

Поскольку уравнения (9.2.39)– (9.2.40) – линейные, то их решение достигается в угловой точке, т.е. задача состоит в том, чтобы из неотрицательных ($X \geq 0$, $Y \geq 0$) базисных решений системы (9.2.39)– (9.2.40) найти такие, которые обращают в нуль произведение $X^T V$. Отсюда приходим к выводу, что если существует оптимальное решение задачи квадратичного программирования, то оно должно быть одним из базисных решений системы уравнений (9.2.39) – (9.2.40). Как было показано выше, оптимальный план задачи линейного программирования также находится среди базисных решений системы ограничений.

Точные алгоритмы решения задачи выпуклого квадратичного программирования преимущественно используют симплекс-метод. Чтобы реализовать эту идею, необходимо решение системы (9.2.43) связать с каким-то критерием отбора базисных решений, так как исследовать все практически не всегда возможно.

9.3. МЕТОД ЛИНЕЙНОЙ АППРОКСИМАЦИИ

Функция n переменных $f(x_1, x_2, ..., x_n)$ называется *сепарабельной*, если ее можно представить как сумму n функций от одной переменной x_j, $j = \overline{1, n}$,

т. е. при условии $f(x_1, x_2, ..., x_n) = \sum_{j=1}^{n} f_j(x_j)$.

Задачи нелинейного программирования, у которых функция цели и функции системы ограничений сепарабельны, называются задачами *сепарабельного программирования*. Таким образом, общая задача сепарабельного программирования примет вид

$$\max(\min) Z = \sum_{j=1}^{n} z_j(x_j) \qquad (9.3.1)$$

при ограничениях:

$$\sum_{j=1}^{n} g_{ij}\left(x_j\right) * b_i \,,\ \overline{1,m}\,; \qquad (9.3.2)$$

$$x_j \geq 0,\ j=\overline{1,n} \qquad (9.3.3)$$

где $*$ – отношение <, > или (=).

Задачи сепарабельного типа можно решать общими методами нелинейного программирования. Однако с учетом специфики разработан специальный приближенный метод, позволяющий свести задачу к линейному программированию. Он основан на *кусочно-линейной аппроксимации* (приближении) функций $z_j(x_j)$ и $g_{ij}(x_j)$.

В курсе вычислительной математики доказывается, что любая непрерывная функция $f(x)$ может быть аппроксимирована с наперед заданной степенью точности кусочно-линейной функцией $\hat{f}(x)$. Этот факт положен в основу линейной аппроксимации задач с сепарабельными функциями. Заменяем $z_j(x_j)$ и $g_{ij}(x_j)$ кусочно - линейными функциями. Тогда для задачи (9.3.1) – (9.3.3) получим ее приближенный аналог

$$\max(\min)\hat{Z} = \sum_{j=1}^{n} \hat{z}_j(x_j) \qquad (9.3.4)$$

при ограничениях:

$$\sum_{j=1}^{n} \hat{g}\left(x_j\right) * b_i \,,\ \overline{1,m}\,; \qquad (9.3.5)$$

$$x_j \geq 0,\ j=\overline{1,n} \qquad (9.3.6)$$

Рассмотрим построение аналитическогое выражения аппроксимирующей кусочно-линейной функции. Пусть задана непрерывная функция одной переменной $h(x)$, определенная на отрезке $[0, a]$. Разобьем отрезок $[0, a]$ на r частей. Пусть точки деления отрезка $x_0=0$, $x_1<x_2<...<x_r=a$. Найдем для них значение функции $h_k=h(x_k)$, $k=\overline{0,r}$. Соединим полученные точки в порядке следования отрезками прямых. Полученная ломаная линия аппроксимирует функцию $h(x)$ на отрезке $[0, a]$. Найдем аналитическое выражение $\hat{h}(x)$ через значения функции $h(x)$ в точках x_k, $k=\overline{0,r}$. Проанализируем две последовательные точки деления $x_к$ и x_{k+1}, значения функции $h(x)$ в них: h_k и h_{k+1} (рис. 9.1). На отрезке $[x_k, x_{k+1}]$ заменим уравнение функции $h(x)$ уравнением прямой линии, проходящей через точки (x_k, h_k), (x_{k+1}, h_{k+1}).

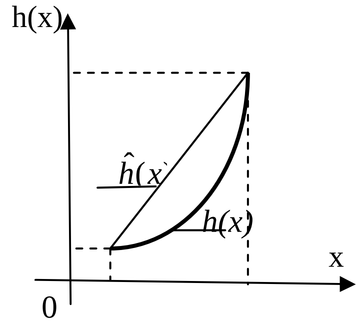


Рис.9.1

Для всего промежутка *[0, a]* получим:

$$x = \sum_{k=0}^{r} \lambda_k x_k \; ; \qquad (9.3.7)$$

$$\hat{h}(x) = \sum_{k=0}^{r} \lambda_k \hat{h}_k \; ; \qquad (9.3.8)$$

$$\sum_{k=0}^{r} \lambda_k = 1, \quad \lambda_k \geq 0, \, k = \overline{0, r} \, ; \qquad (9.3.9)$$

Очевидно, при таком представлении отличными от нуля могут быть только одно или два соседних λ_k. Выражение (9.3.8) есть аналитическая запись аппроксимирующей кусочно-линейной ломаной линии. Используя формулы (9.3.7)-(9.3.9), запишем аналитические выражения для аппроксимирующих функций (9.3.4)-(9.3.5). Определим максимальное значение a_j, которое может принимать переменная x_j. Разобьем каждый из отрезков $[0, a_j]$ на r_j частей с помощью точек x_{kj}, где $x_{0j} = 0$, $x_{ri} = a_j$. Тогда функции Z и g_{ij} задачи (1) – (3) примут вид

$$\hat{Z} = \sum_{j=1}^{n} \sum_{k=0}^{r} z_{kj} \lambda_{kj}, \quad \hat{g}_{ij} = \sum_{j=1}^{n} \sum_{k=0}^{r_j} g_{kij} \lambda_{kj}$$

Задача (9.3.1)– (9.3.3) запишется так:

$$\max(\min) : \hat{Z} = \sum_{j=1}^{n} \sum_{k=0}^{r} z_{kj} \lambda_{kj}, \qquad (9.3.10)$$

при ограничениях:

$$\sum_{j=1}^{n} \sum_{k=0}^{r_j} g_{kij} \lambda_{kj} \, \square \, b_i, \; i = \overline{1, m} \qquad (9.3.11)$$

$$\sum_{k=0}^{r_j} \lambda_{kj} = 1, \quad j = \overline{0, n} \, ; \qquad (9.3.12)$$

$$\lambda_{kj} \geq 0\,, k=\overline{0,r_j}\,, \quad j=\overline{1,n}\,; \tag{9.3.13}$$

Задача (9.3.10)-(9.3.13) относительно переменных λ_{kj} называется аппроксимирующей для (9.3.1)-(9.3.3). Если $\lambda_{kj}{}^{*}$– оптимальное решение аппроксимирующей задачи (9.3.10)– (9.3.13), тогда оптимальное решение исходной можно найти по формулам

$$x_j^* = \sum_{k=0}^{r_j} \lambda_{kj}^* x_{kj}\,, \quad j=\overline{1,n} \tag{9.3.14}$$

Если некоторая переменная x_j входит в задачу линейно, то ее не следует выражать через λ_{kj}, а использовать как переменную в аппроксимирующей задаче.

9.4. ПРИБЛИЖЕННЫЕ МЕТОДЫ НЕЛИНЕЙНОЙ ОПТИМИЗАЦИИ

9.4.1. Метод релаксации Зейдел.

Относится к приближенным методам нелинейной оптимизации.

Релаксационные методы направлены на уменьшение трудностей, связанных с отысканием экстремума функции цели со сложной аналитической структурой классическими методами дифференциального исчисления. Рассмотрим стандартный релаксационный прием отыскания экстремумов нелинейной функции многих переменных.

Пусть задана непрерывная дифференцируемая функция многих переменных $f(x_1, x_2, ..., x_n)$. Следует найти ее экстремальное значение. Выберем произвольную точку в области определения функции

$$X^0 = (x_1^0, x_2^0, ..., x_n^0)\,.$$

Зафиксировав значения всех ее координат, кроме одной, будем по этой координате находить экстремальное значение целевой функции. Например, зафиксировав значения переменных $x_2^0, x_3^0, ..., x_n^0$, будем изменять координату x_1 и отыскивать экстремум функции одной переменной $f(x_1, x_2^0, ..., x_n^0)$. Для этого решим уравнение

$$\frac{\partial f(x_1, x_2^0, ..., x_n^0)}{\partial x_1} = 0\,.$$

Исследовав его стационарные решения, найдем экстремум функции $f(x_1, x_2^0, ..., x_n^0)$. Пусть он достигается в точке $x_1 = x_1^{(1)}$. Возьмем новую точку $x_1^{(1)}, x_2^0, ..., x_n^0$. Зафиксируем ее координаты $x_1^{(1)}, x_3^0, ..., x_n^0$ и будем изменять координату x_2. Находим экстремум функции одной переменной $f(x_1^{(1)}, x_2, x_3^0, ..., x_n^0)$. Для этого решим уравнение

$$\frac{\partial f(x_1^{(1)}, x_2, x_3^0 ..., x_n^0)}{\partial x_2} = 0 .$$

Пусть экстремум этой функции достигается в точке $x_2 = x_2^{(1)}$. Далее находим экстремум функции $f(x_1^{(1)}, x_2^1, x_3, x_4^0, ..., x_n^0)$ и т. д. Этот прием повторяется *n* раз. В результате переходим к точке $x_1^{(1)}, x_2^{(1)}, ..., x_n^{(1)}$. Вновь фиксируем координаты $x_2^{(1)}, x_3^{(1)}, ..., x_n^{(1)}$ и рассматриваем функцию одной переменной $f(x_1, x_2^{(1)}, x_3^{(1)}, x_4^{(1)}, ..., x_n^{(1)})$. Пусть решением уравнения

$$\frac{\partial f(x_1, x_2^{(1)}, ..., x_n^{(1)})}{\partial x_1} = 0$$

является точка $x_1 = x_1^{(2)}$ и т. д.

Признаком оптимальности плана служит равенство нулю первых частных производных функций *f(X)* в точке, полученной в результате последнего релаксационного шага. Очевидно, если целевая функция аддитив-на, то экстремум будет достигнут за один цикл. Метод прост в реализации, однако сходимость процедуры крайне чувствительна к ориентации линий уровня оптимизируемой функции относительно осей координат. Например, пусть отыскивается минимум функции двух переменных

$$f(x_1, x_2) = a^2 x_1^2 + b^2 x_2^2 .$$

Линии уровня целевой функции

$$a^2 x_1^2 + b_2 x_2^2 = const$$

представляют собой эллипсы, оси которых ориентированы по координатным осям. Если выбрать в качестве начальной точки произвольную точку *A* (рис. 9.2), то релаксационная процедура приведет к минимуму

$$x_1^* = 0, x_2^* = 0$$

за цикл из двух шагов (рис. 9.3). Повернем оси эллипса на 45°. Вместо поворота осей эллипса можно повернуть оси координат на 45° (рис. 9.4).

Теперь для достижения минимума целевой функции теоретически необходимо сделать бесконечное число циклов. После каждого переходим в точки A_1, A_2 и т.д. (Рис.9.4). Это снижает эффективность релаксационной процедуры. Другой недостаток метода связан с необходимостью иметь аналитические выражения для частных производных целевой функции и с трудностями решения уравнений

$$\frac{\partial f(X)}{\partial x_j} = 0, \quad j = \overline{1,n}$$

Продвижение к экстремуму будет более эффективным, если выбор переменной оптимизации осуществить следующим образом. Выбирается произвольная точка A_o, принадлежащая области определения функции. Определяются частные производные

$$\frac{\partial f(A_0)}{\partial x_j}, \quad j = \overline{1,n}.$$

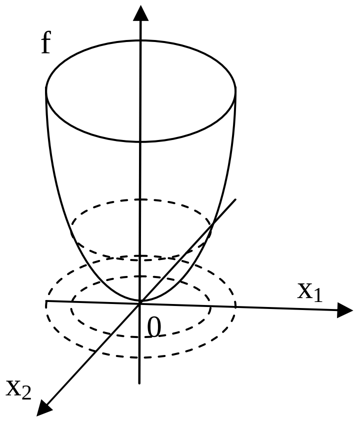


Рис.9.2

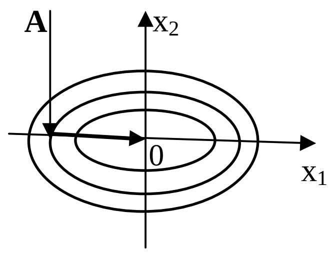


Рис.9.3

Находим наибольшую по абсолютной величине производную целеевой функции в начальной точке A_0. Пусть

$$\max_j \left| \frac{\partial f(A_0)}{\partial x_j} \right| = \left| \frac{\partial f(A_0)}{\partial x_{j_1}} \right|.$$

Фиксируем все координаты точки A_0, кроме x_{j_1}. Находим экстремум функции *f(X)* по переменной x_{j_1} при фиксированных значениях остальных переменных, определяемых точкой A_0. Пусть экстремум функции по x_{j_2} достигается в точке $x_{j_1} = x_{j_1}^1$. Переходим к точке

$$(x_1^0, ..., x_{j_1-1}^0, x_{j_1}^1, x_{j_1+1}^0, ..., x_{j_2-1}^0, x_{j_2}^1, x_{j_2+1}^0, ..., x_n^0)$$

и т.д. Критерий окончания процедуры – равенство нулю всех частных производных целевой функции. Практически, когда процедура может иметь бесчисленное множество циклов, в качестве критерия окончания принимается условие $\sum_{j=1}^{n} \left(\frac{\partial f(X)}{\partial x_j} \right)^2 \le \varepsilon$, где $\varepsilon > 0$ – некоторое наперед заданное малое число.

Недостатки методов релаксации преодолеваются при применении градиентных методов.

9.4.2. Градиентные методы

Их сущность вытекает из самого названия. Латинское слово градиент означает меру возрастания или убывания какой-либо величины. В математике градиент–это вектор, показывающий направление наискорейшего возрастания функции. Чтобы понять сущность названных методов, вспомним точное определение градиента и его свойства. Для наглядности все рассуждения проведем на примере целевой функции двух переменных $Z=z(x_l,x_2)$.

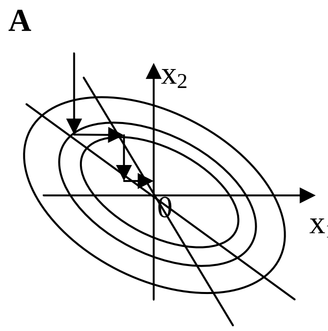


Рис 9.4

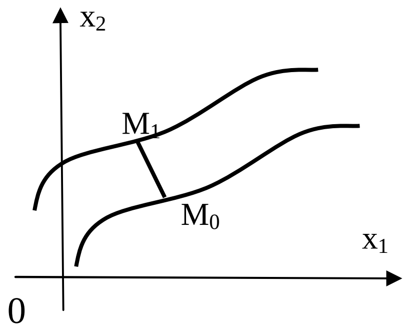


Рис 9.5

Выясним, как изменится значение целевой функции при переходе от одной точки пространства к другой. Для этого рассмотрим какую-либо точку M_0 $\left(x_1^0, x_2^0\right)$ и через нее проведем линию уровня $z(x_1,x_2) = C$, где $C = z\left(x_1^0, x_2^0\right)$. Затем построим соседнюю линию уровня, определяемую уравнением $z(x_1,x_2) = C+\Delta C$. Пусть точка M_1 принадлежит линии уровня $z(x_1,x_2) = C+\Delta C$. Приращение функции ΔZ при переходе из точки M_0 в точку M_1 равно $\Delta Z = \Delta C$. Это приращение не зависит от положения точки M_1 на соседней линии уровня (рис 95). Отношение приращения функции к длине соответствующего отрезка MoM_1 назовем средней скоростью изменения функции $z(x_1,x_2)$ на отрезке M_0M_1. Эта средняя скорость $\Delta Z / M_0M_1$ зависит от длины M_0M_1 иначе говоря, от положения точки M_1 на поверхности уровня $z(x_1,x_2) = C+\Delta C$. Она достигает наибольшей величины по направлению, где длина отрезка M_0M_1 будет наименьшей. Наименьшей же длина отрезка M_0M_1 будет в том случае если направление M_0M_1 перпендикулярно линии уровня $z(x_1,x_2) = C$ при ΔC стремящемся к нулю. Но при этом и длина отрезка будет стремиться к нулю.

Рассмотрим предел отношения приращения функции ΔZ к длине M_0M_1, когда длина отрезка стремится к нулю. Величина предела называется скоростью изменения функции $z(x_1,x_2)$ в точке M_0 по направлению вектора M_0M_1 . Как следует из выше сказанного, скорость изменения функции в точке M_0 будет наибольшей, если направление M_0M перпендикулярно касательной к кривой $z(x_1,x_2)= C$ в точке M_0. Как известно, физический смысл производной в точке M есть скорость изменения функции $z(X)$ по направлению оси x. Следовательно, частные производные функции $z(x_1,x_2)$ по x_1, x_2 в произвольной точке M_0 – это скорости изменения функции по координатным осям. Обозначим единичные векторы (орты) по координатным осям через $\vec{e}_1, \vec{e}_2$. Тогда вектор максимального изменения целевой функции в точке M_0 примет вид

$$\vec{n} = \frac{\partial z(M_0)}{\partial x_1}\vec{e}_1 + \frac{\partial z(M_0)}{\partial x_2}\vec{e}_2 . \qquad (9.4.1)$$

Длина этого вектора равна

$$|\vec{n}| = \sqrt{\left(\frac{\partial z(M_0)}{\partial x_1}\right)^2 + \left(\frac{\partial z(M_0)}{\partial x_2}\right)^2} \qquad (9.4.2)$$

Из формулы (9.4.2) следует, что в точке M_0 вектор наибольшего изменения целевой функции равен нулю, когда

$$\frac{\partial z(M_0)}{\partial x_1} = 0, \frac{\partial z(M_0)}{\partial x_2} = 0 ,$$

а это значит, что функция $z(x_1,x_2)$ – постоянная величина. Вектор $\vec{n}$, определяемый формулой (9.4.1), называется градиентом функции в точке M_0. Градиент функции $z(x_1,x_2)$ в точке M обозначается так: *grad* $z(M_0)$ или $\nabla z(M_0)$. Обобщая сказанное на случай произвольной целевой функции, придем к определению: *grad Z(X) есть вектор, направленный по нормали к поверхности уровня целевой функции в сторону ее возрастания по этому направлению*. Чтобы получить скорость изменения функции Z по любому направлению, достаточно спроектировать *gradZ* на это направление.

Итак, для функции n переменных $z(x_1,x_2, ..., x_n)$

$$\overrightarrow{grad}\, z(X) = \frac{\partial z}{\partial x_1}\vec{e}_1 + \frac{\partial z}{\partial x_2}\vec{e}_2 + ... + \frac{\partial z}{\partial x_n}\vec{e}_n \qquad (9.4.3)$$

где $\vec{e}_1, \vec{e}_2, ..., \vec{e}_n$ – орты. Вместо (9.4.39) градиент часто записывают его проекциями на координатные оси, т.е. так

$$\overrightarrow{grad}\, z(X)=\left(\frac{\partial z}{\partial x_1},\frac{\partial z}{\partial x_2},...,\frac{\partial z}{\partial x_n}\right)$$

Вектор [*-grad* $z(M_0)$] называется антиградиентом функции $z(X)$ в точке M_0. Он указывает направление наискорейшего убывания функции $z(X)$ в точке M_0.

Так как аналитические выражения для частных производных часто оказываются сложными или вообще отсутствуют, то их вычисляют приближенно так: $\frac{\partial z(X)}{\partial x_j}\approx\frac{\Delta z(X)}{\Delta x_j}=\frac{z(x_1,...,x_j+\Delta x_j,...,x_n)-z(x_1,...,x_j,...,x_n)}{\Delta x_j}$ где Δx_j – приращение координаты x_j.

Сущность градиентных методов заключается в следующем. Пусть, например, находим *max (min)* $z(X)$. Выбираем произвольную начальную точку M_0, принадлежащую области определения функции. 1) Выясняется направление градиента функции в точке M_0. 2) Осуществляется перемещение из точки M_0 на величину, равную некоторому шагу, в точку M_1. Перемещение ведется в направлении градиента, если ищется *max* функции, и в направлении антиградиента, если ищется *min* функции. 3) В точке M_1 устанавливается новое градиентное направление и опять осуществляется перемещение на некоторый шаг в новую точку и т.д.

Как известно, параметрическое уравнение прямой, проходящей через данную точку M в заданном направлении *grad* $z(M)$, имеет вид

$$x_j=x_j(M)+\alpha\frac{\partial z(M)}{\partial x_j},\quad j=\overline{1,n}\,.$$

Шаг перемещения равен величине параметра α. Если мы отыскиваем максимум функции, то должны перемещаться в направлении градиента (в этом случае берется $\alpha>0$); если же определяется минимум – то движение идет в направлении, противоположном градиенту, и берется $\alpha<0$.

Таким образом, в градиентных методах изменяется значение не одной, а всех независимых переменных. Каждая получает приращение, пропорциональное соответствующей составляющей градиента по данной оси. Разновидности градиентных методов отличаются лишь выбором шага вдоль градиентного или антиградиентного направления. Признаком достижения экстремума служит обращение в нуль градиента функции $gradZ=0$. При решении практических задач, как правило, ограничиваются приближенным значением оптимального плана. Критерием окончания итеративной процедуры является

$$\sum_{j=1}^{n}\left(\frac{\partial z(X)}{\partial x_j}\right)^2 \le \varepsilon$$

где ε>0 – наперед заданное число. Существуют и другие критерии окончания поиска. Например, после каждой итерации можно сравнивать достигнутое значение целевой функции со значением в предыдущей точке. Если после очередного шага изменение целевой функции не превышает некоторого наперед заданного числа $\varepsilon>0$, то поиск прекращается.

Рассмотрим простейшие градиентные методы: а) градиентного спуска (подъема) б) наискорейшего спуска (подъема).

Метод градиентного спуска (подъема). Для задачи минимизации говорят о спуске, для задачи максимизации – о подъеме. В этом методе итерации строятся по формуле

$$x_j^{k+1} = x_j^k \pm \alpha\, grad\, z(X^k),$$

где α > 0, знак (-) относится к задачам на *min*; знак (+) – на max; x_j^0 – некоторое начальное приближение. Иначе говоря, в методе градиентного спуска (подъема) на каждой итерации вычисляется направление градиента и делается шаг постоянной длины α в антиградиентном направлении (для задачи на min) или вдоль этого направления (для задачи на max).

Пример. Снабженческая организация поставляет два вида продукции – x_1 и x_2. Издержки обращения, связанные с этим, выражаются функцией потерь

$$z(x_1,x_2) = x_1^2 - 7x_1 + x_2^2 - 4x_2 - x_1x_2 + 35,$$

где x_1 и x_2 измеряются в тыс. штук, а Z – в тыс. руб. Определить план товарооборота, минимизирующий издержки обращения. Вычисления прекратить, как только значения функции цели на двух последовательных итерациях будут отличаться не больше, чем на $0{,}1$ $(\varepsilon=0{,}1)$.

Возьмем в качестве начальной точки $M_0\ (1,\ 1),\ z(M_0) = 25,$ зададимся шагом оптимизации $\alpha=1/2$.

Находим: $\dfrac{\partial Z}{\partial x_1} = 2x_1 - x_2 - 7;\quad \dfrac{\partial Z}{\partial x_2} = 2x_2 - x_1 - 4;$

$$\frac{\partial z(M_0)}{\partial x_1} = -6; \quad \frac{\partial z(M_0)}{\partial x_2} = -3.$$

Переходим к новой точке $M_1(x_1^1, x_2^1)$:

$$\begin{cases} x_1^1 = 1 - \frac{1}{2}(-6) = 4, \quad \frac{\partial z(M_1)}{\partial x_1} = -\frac{3}{2}, \quad \frac{\partial z(M_1)}{\partial x_2} = -3, \\ x_2^1 = 1 - \frac{1}{2}(-3) = \frac{5}{2}, \qquad z(M_1) = 9,25. \end{cases}$$

Так как *z(M$_0$)- z(M$_1$)=25-9,25>ε,* то план не оптимальный.

Аналогично переходим к точкам *M$_2$, M$_3$, M$_4$, M$_5$:*

$$M_2(x_1^2, x_2^2): \begin{cases} x_1^2 = 4 - \frac{1}{2}\left(-\frac{3}{2}\right) = \frac{19}{4}, \\ x_2^2 = \frac{5}{2} - \frac{1}{2}(-3) = 4. \end{cases}$$

$$\frac{\partial z(M_2)}{\partial x_1} = -\frac{3}{2}, \quad \frac{\partial z(M_2)}{\partial x_2} = -\frac{3}{4}, \quad z(M_2) \approx 5,31;$$

$$z(M_1) - z(M_2) = 9,25 - 5,31 > \varepsilon;$$

$$M_3(x_1^3, x_2^3) = M_3\left(\frac{11}{2}; \frac{35}{8}\right), \quad z(M_3) \approx 4,33;$$

$$z(M_2) - z(M_3) = 5,31 - 4,33 > \varepsilon;$$

$$M_4(x_1^4, x_2^4) = M_4\left(\frac{91}{16}; \frac{19}{4}\right), \quad z(M_4) \approx 4,08;$$

$$z(M_3) - z(M_4) = 4,33 - 4,08 > \varepsilon;$$

$$M_5(x_1^5, x_2^5) = M_5\left(\frac{47}{8}; \frac{155}{32}\right), \quad z(M_5) \approx 4,02;$$

$$z(M_4) - z(M_5) = 4,08 - 4,02 < \varepsilon = 0,1;$$

$$x_1^* = \frac{47}{8} \approx 5,88; \; x_2^* = \frac{155}{32} \approx 4,84; \; z(x_1^*, x_2^*) \approx 4,02.$$

Сравним полученное решение с точным, которое получается путём решения системы

$$\begin{cases} \frac{\partial Z}{\partial x_1} = 2x_1 - x_2 - 7 \\ \frac{\partial Z}{\partial x_2} = 2x_2 - x_1 - 4 \end{cases} \quad \textit{откуда} \quad x_1^* = 6, \; x_2^* = 5,, \; Z\left(x_1^*, x_2^*\right) = 4.$$

Итак, оптимальным планом товарооборота будет: $x_1 \approx 5{,}88$ тыс. шт. $x_2 \approx 4{,}84$ тыс. шт. При этом минимальные издержки составят *4,02* тыс. руб.

В заключение отметим, что выбор длины шага – задача важная и достаточно сложная. Если длина шага большая, то есть возможность перескочить экстремум. При этом появляется опасность *зацикливания.* Если длина шага малая, то процесс сходимости к экстремальной точке может оказаться очень медленным. Трудности, связанные с выбором длины шага, преодолеваются в так называемом методе наискорейшего спуска (подъема), предложенном Коши.

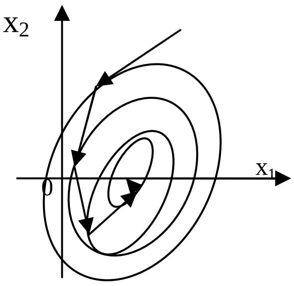


Рис.9.6

Метод наискорейшего подъема (спуска). В этом методе движение вдоль градиентного направления производится до тех пор, пока функция продолжает возрастать (для задачи на *max*), или в направлении антиградиента, пока функция продолжает убывать (для задачи на *min*) (рис. 9.6). Пусть отыскивается max целевой функции. Выбирается начальная точка M_0. Находится градиент функции в точке M_0: $grad z(M_0)$. В направлении градиента делается шаг. Если значение целевой функции в результате шага увеличилось, то делается еще один шаг и т. д. Движение вдоль этого же направления идет до тех пор, пока функция не перестанет возрастать. Находим точку, в которой в заданном направлении функция достигает max. Пусть это будет точка M_1. Для нее вычисляем новое градиентное направление grad $z(M_1)$. Вдоль него перемещаемся до тех пор, пока целевая функция не перестанет возрастать, и т.д.

В общем виде градиентную процедуру наискорейшего спуска (подъема) можно представить так:

$$X^{k+1} = X^k \pm \alpha_k grad\, z(X^k). \qquad (9.4.4)$$

Здесь величина шага α_k выбирается из условия

$$z\left(X^k \pm \alpha_k grad\, z(X^k)\right) = \max_{\alpha>>0}(\min) z\left(X^k \pm \alpha grad\, z(X^k)\right), \qquad (9.4.5)$$

где max и знак (+) относятся к задаче максимизации, а min и знак (-) – к задаче минимизации. Условие (9.4.5) означает, что движение вдоль градиентного направления происходит до тех пор, пока функция *z(X)* продолжает возрастать (для задачи на *max*) или убывать (для задачи на *min*).

Пример. Найдем методом наискорейшего спуска минимум функций потерь, рассмотренной в предыдущем пункте. Начальная точка $M_0(1,\ 1)$. Воспользуемся условием (9.4.5) для определения величины шага движения из точки M_0 в антиградиентном направлении:

$$x_1^1(\alpha_1) = x_1(M_0) - \alpha_1 \frac{\partial z(M_0)}{\partial x_1} = 1 - \alpha_1(-6) = 1 + 6\alpha_1;$$

$$x_2^1(\alpha_1) = x_2(M_0) - \alpha_1 \frac{\partial z(M_0)}{\partial x_2} = 1 - \alpha_1(-3) = 1 + 3\alpha_1.$$

Подставим эти выражения в формулу для $z(x_1,x_2)$:

$z(\alpha_1)=(1+6\alpha_1)^2-7(1+6\alpha_1)+(1+3\ \alpha_1)^2-4(1+3\ \alpha_1)-(1+6\ \alpha_1)(1+3\ \alpha_1)+35.$

После раскрытия скобок и приведения подобных членов получим

$$z(\alpha_1)=27\ \alpha_1^2-45\ \alpha_1+35.$$

Найдем α_1 при котором функция $z(\alpha_1)$=достигает *min*. Это, очевидно, α_1 = l/6.

Итак, в результате первого шага перейдем в точку $M_1(x_1^1, x_2^1)$:

$$x_1^1 = 1 - \frac{5}{6}(-6) = 6, \quad \frac{\partial z(M_1)}{\partial x_1} = \frac{3}{2}, \quad x_2^1 = 1 - \frac{5}{6}(-3) = \frac{7}{2}; \quad \frac{\partial z(M_1)}{\partial x_2} = -3.$$

Сделаем новый шаг a_2 движения из точки M_1 вдоль $(-gradz(M_1))$

$$x_1^2(\alpha_2) = x_1(M_1) - \alpha_2 \frac{\partial z(M_0)}{\partial x_1} = 6 - \frac{3}{2}\alpha_2;$$

$$x_2^2(\alpha_2) = x_2(M_1) - \alpha_2 \frac{\partial z(M_1)}{\partial x_2} = \frac{7}{2} + 3\alpha_2.$$

Подставляя $x_1^2(\alpha_2)$ и $x_2^2(\alpha_2)$ в $z(x_1,\ x_2)$, получим $z(\alpha_2) = \frac{3}{4}\left(21\alpha_2^2 - 15\alpha_2 + 7\right)$, откуда $\min\limits_{\alpha_2>0} z(\alpha_2)$ достигается в точке α_2 = 5/14. В результате второго шага перейдем в точку M_2:

$$M_2(x_1^2, x_2^2) = M_2\left(\frac{153}{28}, \frac{32}{7}\right); \quad \frac{\partial z(M_2)}{\partial x_1} = -\frac{9}{14}, \quad \frac{\partial z(M_2)}{\partial x_2} = -\frac{9}{28}, \quad z(M_2)=4{,}24.$$

Сделаем еще один шаг вдоль *(-gradz(*M_2*))*. Найдем его величину, при которой вдоль *(-gradz(*M_2*))* функция цели достигает *min*. Имеем:

$$x_1^3 = \frac{153}{28} + \frac{19}{4}\alpha_3, \quad x_2^3 = \frac{32}{7} + \frac{9}{28}\alpha_3.$$

Подставив эти значения в $z(x_1, x_2)$, найдем $\alpha_3^* = \frac{5}{6}$. Переходим в точку M_3:

$$x_1^3 = \frac{153}{28} + \frac{19}{4}\cdot\frac{5}{6} = 6, \quad x_2^3 = \frac{32}{7} + \frac{9}{28}\cdot\frac{5}{6} \approx 4{,}87; \quad z(M_3) = 4{,}01.$$

Итак, методом наискорейшего спуска получим план более близкий к оптимальному только за три итерации. Градиентные методы можно также применять при отыскании экстремумов нелинейной целевой функции при ограничениях.

РАЗДЕЛ 10. ПРИМЕНЕНИЕ ФУНКЦИЙ, МЕТОДОВ ДИФФЕРЕНЦИАЛЬНОГО И ИНТЕГРАЛЬНОГО ИСЧИСЛЕНИЙ В ЭКОНОМИКЕ

10.1. ПРИМЕНЕНИЕ ФУНКЦИЙ В ЭКОНОМИКЕ

Функции находят широкое применение в экономической теории и практике. Спектр используемых в экономике функций весьма широк: от простейших линейных до функций, получаемый по определенному алгоритму с помощью так называемых рекуррентных соотношений, связывающих состояния изучаемых объектов в разные периоды времени.

Наряду с линейными, используются нелинейные функции, такие, как дробно-линейные (гиперболические), степенные (квадратная, кубическая и т.д.), показательные (экспоненциальные), логарифмические и другие функции. Периодичность ряда экономических процессов позволяет также использовать тригонометрические функции.

Наиболее часто используются в экономике следующие функции:

1. *Функция полезности (функция предпочтений)* – в широком смысле зависимость полезности, т.е. результата, эффекта некоторого действия от уровня (интенсивности) этого действия.
2. *Производственная функция* – зависимость результата производственной деятельности от обусловивших его факторов.
3. *Функция выпуска* (частный вид производственной функции) – зависимость объема производства от наличия или потребления ресурсов.
4. *Функция издержек* (частный вид производственной функции) – зависимость издержек производства от объема продукции.
5. *Функция спроса, потребления и предложения* – зависимость объема спроса, потребления или предложения на отдельные товары или услуги от различных факторов (например, цены, дохода и т.п.).

Если действием побочных факторов можно пренебречь, или удается зафиксировать эти факторы на определенных уровнях, то влияние одного главного фактора изучается с помощью функции одной переменной.

Пример 1. Функции потребления и линия бюджетного ограничения.

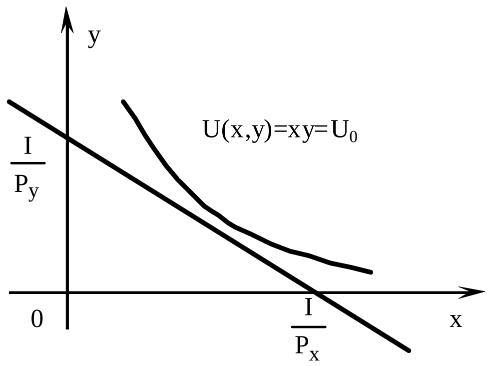


Рис. 10.1

В теории потребительского спроса на два блага x и y (к примеру, исследуемое x и все остальные y), предпочтения потребителя описываются кривой безразличия $U(x,y)=U_k=\text{const}$, а бюджетное ограничение (расходы потребителя ≤ его дохода) в случае, когда потребитель тратит весь свой доход на рассматриваемые блага: $xp_x+yp_y=I$, где I – доход потребителя, а p_x и p_y – цены благ x и y соответственно. Для того, чтобы построить графики этих неявно заданных функций y(x) в системе координат, где по оси абсцисс отложена величина блага x, а по оси ординат – y, нужно выразить в явном виде величину y как функцию x для обеих зависимостей. Сделаем это для простейшей функции полезности U(x,y)=xy. Для уровня полезности (благосостояния) U_0 и дохода I получаем следующие функции: $y=\dfrac{U_0}{x},\ y=\dfrac{I}{p_y}-\dfrac{p_x}{p_y}\cdot x.$ Графиком первой из этих функций (кривой без-различия) является гипербола, а графиком второй (бюджетного ограничения) – прямая линия, имеющая отрицательный наклон, равный по абсолют-ной величине относительной цене блага x и точку пересечения с осью ординат $\left(\dfrac{I}{p_y}\right)$ соответствующую количеству блага y, которое можно приобрести по цене p_x, если потратить на него весь доход I (рис. 10.1).

Пример 2. Кривые спроса и предложения.

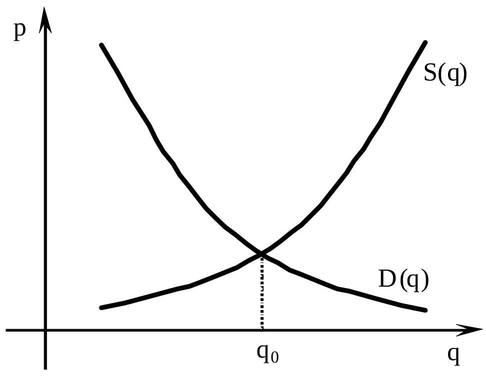


Рис. 10.2

Другим примером функций в экономике служат функции спроса (D) и предложения (S), выражающие связь цены блага (q) и величины спроса и предложения блага при постоянных вкусах потребителей, ценах на другие блага и других параметрах.

Рассматривая в одной системе координат кривые спроса и предложения, можно установить равновесную (рыночную) цену q_0 данного товара в процессе формирования цен в условиях конкурентного рынка (*паутинообразная модель экономики*) (рис. 10.2).

Пример 3. Зависимости величины спроса от дохода.

В модели потребительского спроса используются также функции Л.Торнквиста, моделирующие связь между величиной дохода (I) и величиной спроса потребителей (x) на: а) малоценные товары $\left(x = \frac{\alpha I(I+\beta)}{I^2+\gamma} \right)$; б) товары первой необходимости $\left(x = \frac{\alpha I}{I+\beta} \right)$; в) товары второй необходимости (относительной роскоши) $\left(x = \frac{\alpha(I-\gamma)}{I+\beta} \right)$; г) предметы роскоши $\left(x = \frac{\alpha I(I-\gamma)}{I+\beta} \right)$. Соответствующие им графики приведены на рисунке.

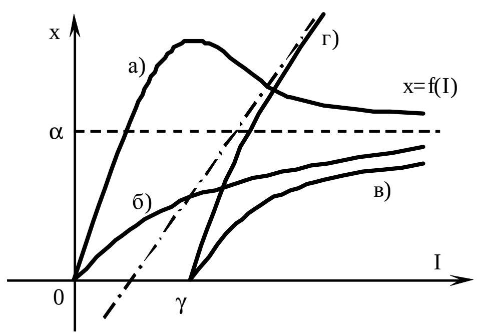


Рис. 10.3

Мы можем установить уровень дохода (γ), при котором начинается приобретение тех или иных товаров и уровень насыщения (α) для групп товаров первой и второй необходимости (рис. 10.3).

Пример 4. Графики зависимости издержек и дохода от объема производства.

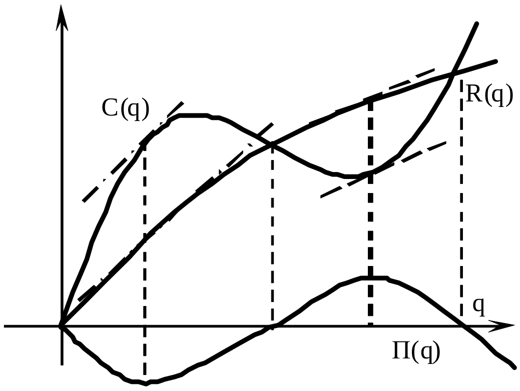


Рис. 10.4

В качестве последнего примера рассмотрим функции издержек C(q) и дохода фирмы $R(q)=q\cdot p(q)$ в зависимости от объема производства q. Поведение функции дохода определяется функцией спроса p(q), рассмотренной выше. Поэтому рассмотрим более подробно поведение функции издержек. В типичном случае издержки фирмы велики при небольшом объеме производства q и вначале растут быстрее, чем доход. С увеличением объема производства скорость роста издержек уменьшается, и в какой-то момент времени они сравниваются с доходом и фирма начинает получать прибыль. При увеличении объема производства прибыль увеличивается, достигая максимума при оптимальном значении q. При дальнейшем увеличении объема производства издержки снова начинают расти быстрее дохода (исчерпаны эффективные ресурсы, нужны дополнительные помещения сырье, квалифицированная рабочая сила) и прибыль фирмы уменьшается, достигая отрицательных значений при достаточно больших объемах производства. Типичные графики дохода, издержек и прибыли приведены на рис. 10.4. Им, например, могут соответствовать функции $R(q)=aq-bq^2$, $C(q)=cq-dq^2+eq^3$.

10.2. ИСПОЛЬЗОВАНИЕ ПОНЯТИЯ ПРОИЗВОДНОЙ В ЭКОНОМИКЕ

Задача о производительности труда. Пусть функция $u=u(t)$ выражает количество произведенной продукции u за время t и необходимо найти производительность труда в момент t_0.

Очевидно, за период времени от t_0 до $t_0+\Delta t$ количество произведенной продукции изменится от значения $u_0=u(t_0)$ до значения $u_0+\Delta u=u(t_0+\Delta t)$;

тогда средняя производительность труда за этот период времени $z_{ср}=\frac{\Delta u}{\Delta t}$.

Производительность труда в момент t_0 можно определить как предельное (маргинальное) значение средней производительности за период времени от t_0 до $t_0+\Delta t$ при $\Delta t \to 0$, т.е.

$$z = \lim_{\Delta t \to 0} z_{cp} = \lim_{\Delta t \to 0} \frac{\Delta u}{\Delta t} = u'_t.$$

Рассмотрим еще одно понятие, иллюстрирующее экономический смысл производной.

Издержки производства y будем рассматривать как функцию количества выпускаемой продукции x. Пусть Δx – прирост продукции, тогда Δy – приращение издержек производства и $\frac{\Delta y}{\Delta x}$ – среднее приращение издержек производства на единицу продукции. Производная $y' = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x}$ выражает предельные (маргинальные) издержки производства и характеризует приближенно дополнительные затраты на производство единицы дополнительной продукции.

Предельные издержки зависят от уровня производства (количества выпускаемой продукции) x и определяются не постоянными производственными затратами, а лишь переменными (на сырье, топливо и т.п.). Аналогичным образом могут быть определены предельная выручка, предельный доход, предельный продукт, предельная полезность и другие *предельные величины*.

Предельные величины характеризуют не состояние (как суммарная или средняя величины), а процесс, изменение экономического объекта. Таким образом, производная выступает как скорость изменения некоторого экономиического объекта (процесса) по времени или относительно другого исследуемого фактора. Следует учесть, однако, что экономика не всегда позволяет использовать предельные величины в силу неделимости многих объектов экономических расчетов и прерывности (дискретности) экономических показателей во времени (например, годовых, квартальных, месячных и т.д.). Вместе с тем в ряде случаев можно отвлечься от дискретности показателей и эффективно использовать предельные величины.

Приведем некоторые обозначения, связанные только с издержками производства (хотя аналогичные вводятся и для других производственных функций), применяемые в экономической теории.

Постоянные издержки (fixed cost, FC) не зависят от объема производства продукции и существуют даже при нулевом его значении (затраты на аренду помещения и оборудования, регистрацию предпринимательской единицы и проч.); *переменные издержки* (variable cost, VC) изменяются пропорционально изменению объема производства (затраты на сырье, материалы и проч.). Сумма постоянных и переменных издержек дает *совокупные* или *валовые издержки* (total cost, TC).

Это деление справедливо только для короткого временного интервала, так как в длинном все издержки становятся переменными.

При расчете на единицу выпускаемой продукции используется понятие *средних издержек* (average cost, AC), причем средние показатели рассчитываются как для *постоянных издержек* (average fixed cost, AFC), так и для *переменных* (average variable cost, AVC) и *валовых* (average total cost, ATC):

$$\mathrm{ATC} = \frac{\mathrm{TC}}{\mathrm{Q}}, \quad \mathrm{AFC} = \frac{\mathrm{FC}}{\mathrm{Q}}, \quad \mathrm{AVC} = \frac{\mathrm{VC}}{\mathrm{Q}}$$

где Q – объем производства продукции.

При изменении объема производства происходит и изменение общего объема издержек. Прирост издержек при увеличении производства продукции на единицу характеризует показатель *предельных издержек* (marginal cost, MC): $\mathrm{MC} = \dfrac{\mathrm{dTC}}{\mathrm{dQ}}$.

Для исследования экономических процессов и решения других прикладных задач часто используется понятие *эластичности функции*.

Эластичностью функции $E_x(y)$ называется предел отношения относительного приращения функции y к относительному приращению переменной x при $\Delta x \to 0$:

$$E_x(y) = \lim_{\Delta x \to 0}\left(\frac{\Delta y}{y} : \frac{\Delta x}{x}\right) = \frac{x}{y}\lim_{\Delta x \to 0}\frac{\Delta y}{\Delta x} = \frac{x}{y} \cdot y'.$$

Эластичность функции показывает приближенно, на сколько процентов изменится функция $y=f(x)$ при изменении независимой переменной x на 1%.

Отметим свойства эластичности функции.

1. Эластичность функции равна произведению независимой переменной x на *темп изменения функции* $T_y = (\ln y)' = \dfrac{y'}{y}$, т.е. $E_x(y)=xT_y$

2. Эластичность произведения (частного) двух функций равна сумме (разности) эластичностей этих функций.

Эластичность функций применяется при анализе спроса и потребления. Например, эластичность спроса y относительно цены x (или дохода x) – коэффициент, показывающий приближенно, на сколько процентов изменится спрос (объем потребления) при изменении цены (или дохода) на 1%.

Если эластичность спроса (по абсолютной величине) $|E_x(y)| > 1$, то спрос считают эластичным, если $|E_x(y)| = 1$ - нейтральным, если $|E_x(y)| < 1$ - неэластичным относительно цены (или дохода). Высокий коэффициент эластичности означает слабое удовлетворение потребностей, низкий - указывает на большую потребность в данном товаре.

10.3. ПРИЛОЖЕНИЕ ПРОИЗВОДНОЙ В ЭКОНОМИЧЕСКОЙ ТЕОРИИ

Рассмотрим некоторые примеры приложения производной в экономической теории. Многие, в том числе базовые, законы теории производства и потребления, спроса и предложения оказываются прямыми следствиями математических теорем.

Вначале рассмотрим экономическую интерпретацию *теоремы Ферма* : если дифференцируемая на промежутке функция достигает наибольшего или наименьшего значения во внутренней точке этого промежутка, то производная функции в этой точке равна нулю.

Один из базовых законов теории производства звучит так: оптимальный для производителя уровень выпуска товара определяется равенством предельных издержек и предельного дохода.

То есть уровень выпуска x_0 является оптимальным для производителя, если $MC(x_0)=MR(x_0)$, где MC – предельные издержки, а MR – предельный доход (marginal revenue).

Обозначим функцию прибыли Z(x). Тогда Z(x)=R(x)-C(x). Очевидно, что оптимальным уровнем производства является тот, при котором прибыль максимальна, т.е. такое значение выпуска x_0, при котором функция Z(x) имеет экстремум (максимум). По теореме Ферма в этой точке $Z'(x_0)=0$. Но $Z'(x)=R'(x)-C'(x)$, поэтому $R'(x_0)=C'(x_0)$ т.е. $MR(x_0)=MC(x_0)$.

Другое важное понятие теории производства – это уровень наиболее экономичного производства, при котором средние издержки по производству товара минимальны. Соответствующий экономический закон гласит: уровень наиболее экономичного производства определяется равенством средних и предельных издержек.

Получим это условие как следствие теоремы Ферма. Средние издержки AC(x) определяются как $\frac{C(x)}{x}$, т.е. издержки по производству товара, деленные на произведенное его количество. Минимум этой величины достигается в критической точке функции y=AC(x), т.е. при условии

$$AC'(x)=\frac{C'x-C}{x^2}=0, \text{ откуда } C'\cdot x-C=0 \text{ или } C'=\frac{C}{x}, \text{ т.е. } MC(x)=AC(x).$$

Понятие выпуклости функции также находит свою интерпретацию в экономической теории.

Один из наиболее знаменитых экономических законов – *закон убывающей доходности* – звучит следующим образом: с увеличением производства дополнительная продукция, полученная на каждую новую единицу ресурса (трудового, технологического и т.д.), с некоторого момента убывает.

Иными словами, величина $\frac{\Delta y}{\Delta x}$, где Δx – приращение ресурса, а Δy – приращение выпуска продукции, уменьшается при увеличении x. Таким образом, закон убывающей доходности формулируется так: функция $y=f(x)$, выражающая зависимость выпуска продукции от вложенного ресурса, является функцией, выпуклой вверх.

Другим базисным понятием экономической теории является функция полезности $U=U(x)$, где x – товар, а U – полезность. Эта величина очень субъективная для каждого отдельного потребителя, но достаточно объектив-ная для общества в целом. *Закон убывающей полезности* звучит следующим образом: с ростом количества товара дополнительная полезность от каждой новой его единицы с некоторого момента убывает. Очевидно, этот закон можно переформулировать так: функция полезности является функцией, выпуклой вверх. В такой постановке закон убывающей полезности служит отправной точкой для математического исследования теории спроса и предложения.

Пример 1. Пусть функция $C(x) = 20x - \frac{x^2}{20}$ устанавливает зависимость издержек производства от количества x выпускаемой продукции. Найти предельные издержки производства и коэффициент эластичности, если объем продукции составляет 100 единиц, 20 единиц.

1. Предельные издержки производства есть производная от функции издержек

$$C'(x) = 20 - \frac{x}{10}.$$

При соответствующих объемах продукции:

$$C'(100) = 20 - \frac{100}{10} = 10;$$

$$C'(20) = 20 - \frac{20}{10} = 18.$$

Итак, чем больше производится продукции, тем медленнее растут издержки на ее выпуск.

2. Эластичность функции $E_x(y) = \frac{x}{y}\frac{dy}{dx}.$

В нашем случае $y(x) = C(x) = 20x - \frac{x^2}{20};$

$$E_x(C) = \frac{x}{20x - \frac{x^2}{20}}\left(20 - \frac{x}{10}\right) = \frac{2(200 - x)}{400 - x};$$

$$E_{100}(C) = \frac{2 \cdot 100}{300} = \frac{2}{3} \approx 0{,}67; \qquad E_{20}(C) = \frac{2 \cdot 180}{380} = \frac{18}{19} \approx 0.95.$$

Итак, при объеме выпуска в 100 единиц, если количество выпускаемой продукции увеличится на 1%, т.е. на 1, то относительные издержки производства увеличатся приближенно на 0,67%; при объеме в 20 единиц увеличение выпуска продукции на 1% повлечет увеличение относительных издержек приближенно на 0,95%.

Пример 2. Зависимость между издержками производства y и объемом выпускаемой продукции x выражается функцией $y=50x-0{,}05x^3$ (ден. ед.). Определить средние и предельные издержки при объеме продукции 10 ед.

Функция средних издержек (на единицу продукции) выражается отношением

$$y_1 = \frac{y}{x} = 50 - 0{,}05x^2,$$

$$y_1(10) = 50 - 0{,}05 \cdot 10^2 = 45 \text{ (ден. ед.).}$$

Предельные издержки: $y' = 50 - 0{,}05 \cdot 3x^2$; $y'(10) = 35$ (ден. ед.).

Итак, если средние издержки на производство единицы продукции составляют 45 ден. ед., то предельные издержки, т.е. дополнительные затраты на производство дополнительной единицы продукции при данном уровне производства (объеме выпускаемой продукции в количестве 10 ед.), составляют 35 ден. ед.

Пример 3. Зависимость между себестоимостью единицы продукции y (тыс. ден. ед.) и выпуском продукции x (млрд. ден. ед.) выражается функцией $y=-0{,}5x+80$. Найти эластичность себестоимости при выпуске продукции равной 60 млрд. ден. ед.

Эластичность себестоимости

$$E_x(y) = \frac{x}{-0{,}5x + 80} \cdot (-0{,}5) = \frac{x}{x - 160},$$

$$E_{60}(y) = \frac{60}{60 - 160} = -0{,}6.$$

Это значит, что при выпуске продукции на 60 млрд. ден. ед., увеличение его на 1% приведет к снижению себестоимости на 0,6%.

Пример 4. Опытным путем установлены функции спроса $D = \dfrac{p+8}{p+2}$ и предложения $S = p + 0{,}5$, где D, S – количество товара соответственно покупаемого и предлагаемого на продажу в единицу времени; p – цена товара. Найти:

1) равновесную цену, т.е. цену, при которой спрос и предложение уравновешиваются;
2) эластичность спроса и предложения для этой цены;

3) изменение дохода при увеличении цены на 5% от равновесной.

Равновесная цена определяется из условия D=S:

$$\frac{p+8}{p+2} = p + 0{,}5,$$

откуда p=2, т.е. равновесная цена равна 2 ден. ед.

Найдем эластичность спроса и предложения по формуле

$$E_x(y) = \frac{x}{y} y'.$$

Тогда $\qquad D' = \dfrac{-6}{(p+2)^2}; \qquad S' = 1;$

$$E_p(D) = -\frac{6p}{(p+2)(p+8)}; \qquad E_p(S) = \frac{2p}{2p+1}.$$

Для равновесной цены p=2 имеем

$$E_2(D) = -0{,}3; \qquad E_2(S) = 0{,}8.$$

Так как полученные значения эластичностей по абсолютной величине меньше единицы, то и спрос, и предложение данного товара при равной (рыночной) цене неэластичны относительно цены. Это означает, что изменение цены не приведет к резкому изменению спроса и предложения.

Так, при увеличении цены p на 1% спрос уменьшается на 0,3%, а предложение увеличивается на 0,8%.

1) При увеличении цены p на 5% от равновесной спрос уменьшается на 5·0,3=1,5%, следовательно, доход возрастает на 3,5% (5%-1,5%=3,5%).

Пример 5. Как связаны предельные и средние затраты предприятия, если эластичность полных затрат равна единице? Пусть полные затраты предприятия y выражаются функцией $y=f(x)$, где x – объем выпускаемой продукции. Тогда средние затраты y_1 на производство единицы продукции $y_1 = \dfrac{y}{x}$. Таким образом, $E_x(y_1) = E_x\left(\dfrac{y}{x}\right) = E_x(y) - E_x(x) = E_x(y) - 1.$

По условию $E_x(y)$=1, следовательно $E_x(y_1)$=1-1=0. Это означает, что с изменением объема продукции x средние затраты на единицу продукции не меняются, т.е. $y_1 = \dfrac{y}{x} = c$, откуда $y=cx$.

Предельные издержки предприятия определяются производной $y' = c$. Итак, $y' = y_1$, т.е. предельные издержки равны средним издержкам. (Заметим, что полученное утверждение справедливо только для линейных функций издержек.)

Пример 6. Фирма планирует выпустить солнечные батареи. На основе исследований была установлена зависимость спроса D от цены p за батарею :

$$D = 100000 - 200p,$$

где D – количество батарей для продажи в год. Затраты фирмы на выпуск D солнечных батарей составляют

$$c = 150000 + 100D + 0{,}003D^2.$$

Рассчитать доход, определить его максимальное значение.

Валовой доход (total revenue)

$$TR = p \cdot D.$$

Из условия запишем функцию p как функцию переменной D:

$$p = 500 - 0{,}005D.$$

Значит, $\quad TR = TR(D) = (500 - 0{,}005D) \cdot D.$

Тогда доход $\quad R = TR - c.$

Итак,

$$R = (500 - 0{,}005D) \cdot D - (150000 + 100D + 0{,}003D^2) = -0{,}008D^2 + 400D - 150000.$$

Определим максимальный доход

$$R'(D) = -0{,}016D + 400,$$

$$-0{,}016D + 400 = 0,$$

$$D = 25000,$$

$$R''(D) = -0{,}016 < 0.$$

Значит, при D=25000 единиц доход достигает максимума, R(25000)=4850000 ден. ед. – максимальное значение дохода.

Пример 7. Предприятие производит x единиц продукции по цене $p(x) = 50 - \frac{1}{10}x$, а издержки производства задаются функцией

$$C(x) = \frac{1}{50}x^2 + 14x + 800.$$

Найти оптимальный для предприятия объем выпуска продукции и соответствующую ему максимальную прибыль.

Пусть TR(x) – валовой доход, z(x) – прибыль от реализации *x* единиц продукции по цене p(x).

Тогда

$$TR(x) = x \cdot p(x);$$
$$z(x) = TR(x) - C(x),$$

где p(x), C(x) – заданные функции.

Для решения задачи следует исследовать функцию z(x) на экстремум. При этом прибыль будет максимальна для такого объема x выпуска продукции, для которого $z'(x) = 0,\ z''(x) < 0.$

Проведем это исследование.

1. Формируем z(x), находим $z'(x)$ и решив уравнение $z'(x) = 0,$ получаем критическую точку.

Учтем, что

$$TR(x) = 50x - \frac{1}{10}x^2, \qquad C(x) = \frac{1}{50}x^2 + 14x + 800;$$
$$z(x) = TR(x) - C(x);$$
$$z'(x) = TR'(x) - C'(x); \qquad TR'(x) = C'(x);$$

$50 - \frac{1}{5}x = \frac{1}{25}x + 14; \qquad \frac{6}{25}x = 36; \qquad x = 150$ – критическая точка.

2. Находим $z''(x)$ и определяем ее знак при x=150:

$$z''(x) = TR''(x) - C''(x) = \frac{1}{5} - \frac{1}{25} = -\frac{6}{25} < 0 \quad \forall x.$$

Значит, x=150 – точка максимума функции z(x), т.е. оптимальный объем производства составляет 150 единиц продукции.

3. Определяем максимальную прибыль производства, т.е. z_{max}=z(150).

При *x=150* цена $p = 50 - \frac{1}{10} \cdot 150 = 35;$ валовой доход *TR=35·150=5250*.

Издержки производства

$$C = \frac{1}{50} \cdot 150^2 + 14 \cdot 150 + 800 = 450 + 2100 + 800 = 3350;$$

максимальная от продажи прибыль z_{max}*=5250-3350=1900*.

Пример 8. Производитель реализует свою продукцию по цене *p* за единицу, а издержки при этом задаются кубической зависимостью $C(x) = ax + \lambda x^3$ $(a < p, \lambda > 0)$. Найти оптимальный для производителя объем выпуска продукции и соответствующую ему прибыль.

Обозначим объем выпускаемой продукции через *x*. Составим функцию прибыли $Z(x) = px - (ax + \lambda x^3)$, где *px* – валовой доход от реализуемой продукции. Исследуем эту функцию на экстремум.

1. Находим $Z'(x) = (p - a) - 3\lambda x^2$.
2. Находим критические точки: $Z'(x) = (p-a) - 3\lambda x^2 = 0$, откуда $x_1 = \sqrt{\dfrac{p-a}{3\lambda}}$ (вторую критическую точку $x_2 = -\sqrt{\dfrac{p-a}{3\lambda}}$ не рассматриваем по смыслу задачи).
3. Находим $Z''(x) = -6\lambda x$ и определяем знак второй производной при $x_1 = \sqrt{\dfrac{p-a}{3\lambda}}$:

$Z''\left(x_1 = \sqrt{\dfrac{p-a}{3\lambda}}\right) < 0$ (в данном случае $Z''(x) < 0$ при любом x>0),

следовательно, при $x = \sqrt{\dfrac{p-a}{3\lambda}}$ прибыль Z(x) максимальна.

4. Находим максимум функции (т.е. максимальный размер прибыли)

$$Z_{max}\left(x_1 = \sqrt{\frac{p-a}{3\lambda}}\right) = \frac{2(p-a)\sqrt{p-a}}{3\sqrt{3\lambda}}.$$

Пример 9. Капитал в 1 млрд. ден. ед. может быть размещен в банке под 50% годовых или инвестирован в производство, причем эффективность вложения ожидается в размере 100%, а издержки задаются квадратичной зависимостью. Прибыль облагается налогом в p%. При каких значениях p вложение в производство является более эффективным, чем чистое размещение капитала в банке?

Пусть x (млрд. ден. ед.) инвестируется в производство, а 1-x размещается под проценты. Тогда размещенный капитал через год станет равным

$$(1-x)\left(1 + \frac{50}{100}\right) = \frac{3}{2} - \frac{3}{2}x,$$

а капитал, вложенный в производство, определим по формуле

$$x\left(1+\frac{100}{100}\right)=2x.$$

Издержки составят ax^2, так как по условию они задаются квадратичной зависимостью, т.е. прибыль от вложения в производство $c=2x-ax^2$.

Налоги составят $(2x-ax^2)\frac{p}{100}$, т.е. чистая прибыль окажется равной $\left(1-\frac{p}{100}\right)(2x-ax^2)$.

Общая сумма через год составит

$$A(x)=\frac{3}{2}-\frac{3}{2}x+\left(1+\frac{p}{100}\right)(2x-ax^2)=\frac{3}{2}+\left[2\left(1-\frac{p}{100}\right)-\frac{3}{2}\right]x-a\left(1-\frac{p}{100}\right)x^2.$$

Требуется найти максимальное значение этой функции на отрезке [0,1]. Имеем

$$A'(x)=2\left(1-\frac{p}{100}\right)-\frac{3}{2}-2a\left(1-\frac{p}{100}\right)x;$$

$$A'(x)=0 \text{ при } x_0=\frac{2\left(1-\frac{p}{100}\right)-\frac{3}{2}}{2a\left(1-\frac{p}{100}\right)};$$

$A''(x)=-2a\left(1-\frac{p}{100}\right)<0$, т.е. x_0 – точка максимума.

Чтобы $x_0\in[0,1]$, необходимо выполнить условие

$$0<2\left(1-\frac{p}{100}\right)-\frac{3}{2}<1 \text{ или } p<25.$$

Таким образом, если p>25, то выгоднее ничего не вкладывать в производство и разместить весь капитал в банк. Если p<25, то можно показать, что при $x=x_0$

$$A(x_0)=\frac{3}{2}+\frac{\left[2\left(1-\frac{p}{100}\right)-\frac{3}{2}\right]^2}{4a\left(1-\frac{p}{100}\right)}>\frac{3}{2}=A(0),$$

т.е. вложение в производство является более выгодным, чем чистое размещение под проценты.

10.4. Экономический смысл определенного интеграла. Использование понятия определенного интеграла в экономике

Пусть функция $z=f(t)$ описывает изменение производительности некоторого производства с течением времени. Найдем объем продукции u, произведенной за промежуток времени $[0,T]$.

Отметим, что если производительность не изменяется с течением времени ($f(t)$ – постоянная функция), то объем продукции Δu, произведенной за некоторый промежуток времени $[t, t+\Delta t]$, задается формулой $\Delta u=f(t)\Delta t$. В общем случае справедливо приближенное равенство $\Delta u \cong f(\xi)\Delta t$, где $\xi \in [t, t+\Delta t]$, которое оказывается тем более точным, чем меньше Δt.

Разобъём отрезок $[0, T]$ на промежутки времени точками:
$0=t_0<t_1<t_2<\ldots<t_n=T$. Для величины объема продукции Δu_i, произведенной за промежуток времени $[t_{i-1}, t_i]$, имеем

$$\Delta u_i=f(\xi_i)\ \Delta t_i,\ где\ \xi_i \in [t_{i-1}, t_i],\ \Delta t_i=t_i-t_{i-1},\ i=1,2,\ldots,n.$$

Тогда $$u \approx \sum_{i=1}^{n} \Delta u_i = \sum_{i=1}^{n} f(\xi)\Delta t_i.$$

При стремлении $\max\limits_i \Delta t_i$ к нулю каждое из использованных приближенных равенств становится все более точным, поэтому

$$u = \lim_{\substack{\max \Delta x_i \to 0 \\ i}} \sum_{i=1}^{n} f(\xi)\Delta t.$$

Учитывая определение определенного интеграла, окончательно получим $u = \int\limits_0^T f(t)dt$, т.е. если $f(t)$ – производительность труда в момент t, то $\int\limits_0^T f(t)dt$ есть объем выпускаемой продукции за промежуток $[0, T]$.

Сравнение данной задачи с задачей о площади криволинейной трапеции показывает, что величина и объем продукции, произведенной за промежуток времени $[0, T]$, численно равен площади под графиком функции $z=f(t)$, описывающей изменение производительности труда с течением времени, на промежутке $[0, T]$ или $\int_0^T f(t)dt$.

Экономический смысл определенного интеграла — объем произведенной продукции при известной функции производительности труда.

Рассмотрим другие примеры использования интеграла в экономике.

Если в функции Кобба-Дугласа* считать, что затраты труда линейно зависят от времени, а затраты капитала неизменны, то она примет вид $g(t)=(\alpha t+\beta)e^{\gamma t}$. Тогда объем выпускаемой продукции за T лет составит:

$$Q = \int_0^T (\alpha t + \beta)e^{\gamma t}dt.$$

Пример 1. Найти объем продукции, произведенной за 4 года, если функция Кобба-Дугласа имеет вид $g(t) = (1+t)e^{3t}$.

По формуле объем Q произведенной продукции равен

$$Q = \int_0^4 (1+t)e^{3t}dt.$$

Используем метод интегрирования по частям. Пусть $u=t+1$, $dv = e^{3t}dt$. Тогда $du = dt, v = \int e^{3t}dt = \frac{1}{3}e^{3t}$.

Следовательно,

$$Q = (t+1)\frac{1}{3}e^{3t}\Big|_0^4 - \int_0^4 \frac{1}{3}e^{3t}dt = \frac{1}{5}(5e^{12}-1) - \frac{1}{9}e^{3t}\Big|_0^4 = \frac{1}{9}(14e^{12}-2) \approx 2{,}53\cdot 10^5 \text{ (усл. ед.).}$$

Исследуя кривую Лоренца – зависимость процента доходов от процента имеющего их населения (кривую ОВА) мы можем оценить степень неравен-

---

* Большинство производственных функций, используемых в экономической теории, являются степенными и имеют вид: $Q=A\cdot L^{\alpha}\cdot K^{\beta}$, где А – коэффициент приведения к одной размерности и учета посторонних факторов, α и β - коэффициенты эластичности выпуска по соответствующим факторам производства, K – объем используемого капитала (оборудование, здания, сооружения), L – объем используемого труда (количество работающих). L и К – переменные факторы производственной функции Q. В общем случае производственная функция может зависеть и от большего количества факторов. Если $\alpha+\beta=1$, то такая функция называется функцией Кобба-Дугласа.

ства в распределении доходов населения. При равномерном распределении доходов кривая Лоренца вырождается в прямую – биссектрису ОА, поэтому площадь фигуры ОАВ между биссектрисой ОА и кривой Лоренца, отнесенная к площади треугольника ОАС (коэффициент Джини), характеризует степень неравенства в распределении доходов населения. Высокое значение этого коэффициента показывает существенно неравномерное распределение доходов среди населения в рассматриваемой стране.

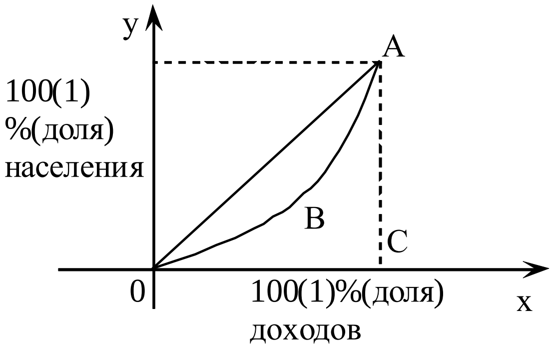


Рис. 10.5

Пример 1. По данным исследований в распределении доходов в одной из стран кривая Лоренца *OBA* может быть описана уравнением $y = a - \sqrt{a^2 - t^2}$, где x – доля населения, y – доля доходов населения. Вычислить коэффициент Джини.

Очевидно, коэффициент Джини

$$k = \frac{S_{OAB}}{S_{\Delta OAC}} = \frac{S_{\Delta OAC} - S_{OBAC}}{S_{\Delta OAC}} \;;$$

$$S_{OBAC} = \int_0^a (a - \sqrt{a^2 - x^2})dx = a\int_0^a dx - \int_0^a \sqrt{a^2 - x^2}dx = a^2 - \int_0^1 \sqrt{a^2 - x^2}dx = a^2 - I \;.$$

$$I = \int_0^a \sqrt{a^2 - x^2}dx \left\| \begin{matrix} x = a\sin t, dx = a\cos t dt, \\ \sqrt{a^2 - x^2} = a\cos t \end{matrix} \right\| = \int_0^{\frac{\pi}{2}} a^2 \cos^2 t dt = \frac{a^2}{2}\int_0^{\frac{\pi}{2}} (1 + cos2t)dt = \frac{\pi a^2}{4} \;.$$

$$S_{OBAC} = a^2 - \frac{\pi a^2}{4} = a^2\left(1 - \frac{\pi}{4}\right);$$

$$S_{\Delta OAC} = \frac{a^2}{2}\;;\quad S_{OBA} = \frac{a^2}{2} - a^2 + \frac{\pi a^2}{4} = \frac{\pi a^2}{4} - \frac{a^2}{2} = \frac{a^2}{2}\left(\frac{\pi}{2} - 1\right);$$

Итак, коэффициент Джини $k = \dfrac{(a^2/2)\left(\dfrac{\pi}{2}-1\right)}{a^2/2} = \dfrac{\pi}{2} - 1 \approx 0,57$

Достаточно высокое значение k показывает неравномерное распределение доходов среди населения в рассматриваемой стране.

Определение начальной суммы по ее конечной величине, полученной через время t (лет) при годовом проценте (процентной ставке) q, называется *дисконтированием.* Задачи такого рода встречаются при определении экономической эффективности капитальных вложений.

Пусть A_t – конечная сумма, полученная за t лет, и A – дисконтируемая (начальная) сумма, которую в финансовом анализе называют также современной суммой. Если проценты простые, то $A_t=A(1+rt)$, где $r=q/100$ – удельная процентная ставка. Тогда $A=A_t/(1+rt)$. В случае сложных процентов $A_t=A(1+rt)^t$ и потому $A=A_t/(1+rt)^t$.

Пусть Q_t – конечная сумма, полученная за T лет, и Q – дисконтируемая (начальная) сумма, которую в финансовом анализе называют также современной суммой. Пусть поступающий ежегодно доход изменяется во времени и описывается функцией $f(t)$ и при удельной норме процента, равной r, процент начисляется непрерывно. Можно показать, что в этом случае дисконтируемая сумма вычисляется по формуле:

$$Q = \int_0^T f(t)e^{-rt}dt.$$

Пример 2. Определить дисконтированный доход за T лет при процентной ставке $r\%$, если первоначальные (базовые) капиталовложения составили b тыс. ден. ед., и намечается непрерывно увеличивать капиталовложения на a тыс. ден. ед.

Очевидно, что капиталовложения задаются функцией $f(t)=b+a\cdot t$. Тогда дисконтируемая сумма капиталовложений $Q = \int_0^T (b+at)e^{-rt}dt.$

Интегрируя по частям, получим $Q = -\dfrac{e^{-rT}}{r^2}\big((a\cdot T+b)\cdot r+a\big)+\dfrac{b\cdot r+a}{r^2}.$

Пусть $T=2,\ b=10,\ a=2,\ r=6\%$,

тогда $\quad Q=-\dfrac{e^{-0,06\cdot 2}}{(0,06)^2}\big((2\cdot 2+10)\cdot 0,06+2\big)+\dfrac{10\cdot 0,06+2}{(0,06)^2}=-246,34\cdot 2,84+722,22=22,6$

Пусть $T=4,\ b=10,\ a=2,\ r=6\%$, тогда

$$Q=-\frac{e^{-0,06\cdot 4}}{(0,06)^2}\big((2\cdot 4+10)\cdot 0,06+2\big)+\frac{10\cdot 0,06+2}{(0,06)^2}=-218,51\cdot 3,08+722,22=49,21$$

Это означает, что для получения одинаковой наращенной суммы через 4 года непрерывные капиталовложения от 10 до 12 тыс. ден. ед. равносильны одновременному первоначальному вложению 49,2 тыс. ден. ед. при той же, начисляемой непрерывно, процентной ставке.

Пример 3. Определить дисконтированный доход за три года при процентной ставке 8%, если первоначальные (базовые) капиталовложения составили 10 млрд. ден. ед., и намечается ежегодно увеличивать капиталовложения на 1 млрд. ден. ед.

Очевидно, что капиталовложения задаются функцией $f(t)=10+1\cdot t=10+t$. Тогда дисконтированная сумма капиталовложений $A=\int\limits_0^3 (10+t)e^{-0,08t}dt.$

Интегрируя по частям, получим А=30,5 млрд. ден. ед. Это означает, что для получения одинаковой наращенной суммы через три года ежегодные капиталовложения от 10 до 13 млрд. ден. ед. равносильны одновременным первоначальным вложениям 30,5 млрд. ден. ед. при той же, начисляемой непрерывно, процентной ставке.

Пусть известна функция $t=t(x)$, описывающая изменение затрат времени t на изготовление изделия в зависимости от степени освоения производства, где x – порядковый номер изделия в партии. Тогда среднее время $t_{ср}$, затраченное на изготовление одного изделия в период освоения от x_1 до x_2 изделий, вычисляется по теореме о среднем:

$$t_{cp}=\frac{1}{x_2-x_1}\int\limits_{x_1}^{x_2} t(x)dx.$$

Что касается функции изменения затрат времени на изготовление изделий $t=t(x)$, то часто она имеет вид

$$t=ax^{-b},$$

где a – затраты времени на первое изделие, b – показатель производственного процесса.

Пример 4. Найти среднее время, затраченное на освоение одного изделия в период освоения от x_1=100 до x_2=121 изделий, полагая в формуле a=600 (мин.), b=0,5.

Используя формулу, получаем

$$t_{cp}=\frac{1}{121-100}\int_{100}^{121}600x^{-1/3}dx=\frac{600}{21}2\sqrt{x}\Bigg|_{100}^{121}=\frac{400}{7}\approx 57{,}2 \text{ (мин.).}$$

Пример 5. Найти среднее значение издержек $f(x)=5x^2+3x+2$, выраженных в денежных единицах, если объем продукции x меняется от 1 до 4 единиц. Указать объем продукции, при котором издержки принимают среднее значение.

Согласно теореме о среднем значении интеграла имеем:

$$f(\xi)=\frac{1}{b-a}\int_a^b f(x)dx.$$

В нашем случае

$$f(\xi)=\frac{1}{3}\int_1^4(5x^2+3x+1)dx=\frac{1}{3}\left(5\frac{x^3}{3}+\frac{3}{2}x^2+2x\right)\Bigg|_1^4=\frac{1}{3}\left((\frac{4^3}{3}-\frac{1}{3})+\frac{3}{2}\cdot(4^2-1)+2\cdot(4-1)\right)=16,5$$

т.е. среднее значение издержек равно *16,5 ден.ед*.

Определим, при каком объёме продукции издержки принимают это значение, т.е. решим уравнение

$$5x^2+3x+2=16{,}5 \text{ или } 10x^2+6x-33=0.$$

$$x_{1,2}=\frac{-6\pm\sqrt{36+1320}}{20}=\frac{-6\pm2\sqrt{339}}{20}=\frac{-3\pm\sqrt{339}}{10}$$

Учитывая, что объем продукции не может быть отрицательным, из последнего уравнения имеем $\xi\approx\frac{-3+18,4}{10}\approx1,5$ т.е. $\xi=1{,}5$ единицы продукции.

10.5. Применение несобственных интегралов в теории вероятностей и экономике

Закон распределения вероятностей непрерывной случайной величины задается с помощью так называемой «дифференциальной функции» или плотности вероятностей – *f(x)*.

Свойства плотности распределения вероятностей:

1. $f(x) \geq 0$;
2. $\int_{-\infty}^{\infty} f(x)dx = 1$, т.е. площадь под кривой плотности распределения равна 1.

Зная плотность распределения, можно найти функцию распределения: $F(x) = \int_{-\infty}^{x} f(x)dx$. Все числовые характеристики непрерывных случайных величин, так называемые, моменты определяются через функцию плотности вероятностей.

Начальный теоретический момент порядка k непрерывной случайной величины X определяется равенством $\nu_k = \int_{-\infty}^{\infty} x^k f(x)dx$.

Центральный теоретический момент k-го порядка непрерывной случайной величины X определяется равенством $\mu_k = \int_{-\infty}^{\infty} \left(x - M(x)\right)^k f(x)dx$, где $M(x) = \nu_1 = \int_{-\infty}^{\infty} x \cdot f(x)dx$ – начальный момент 1-го порядка, который называется математическим ожиданием случайной величины.

Пример 1. Плотность распределения непрерывной случайной величины X задана на всей оси *OX* равенством $f(x)=C/(1+x^2)$. Найти постоянный множитель C.

Используя свойство плотности распределения, напишем:

$$\int_{-\infty}^{\infty} \frac{C}{1+x^2}dx = 1, \quad \text{отсюда} \quad C = \frac{1}{\int_{-\infty}^{\infty} \frac{dx}{1+x^2}};$$

$$\int_{-\infty}^{\infty} \frac{1}{1+x^2}dx = \int_{-\infty}^{0} \frac{dx}{1+x^2} + \int_{0}^{\infty} \frac{dx}{1+x^2} = \lim_{A\to-\infty} \int_{A}^{0} \frac{dx}{1+x^2} + \lim_{B\to\infty} \int_{0}^{B} \frac{dx}{1+x^2} =$$

$$= \lim_{A\to-\infty} arctgx\Big|_A^o + \lim_{B\to\infty} arctgx\Big|_0^B = \lim_{A\to-\infty}(arctg0 - arctgA) + \lim_{B\to\infty}(arctgB - arctg0) = \frac{\pi}{2} + \frac{\pi}{2} = \pi .$$

Ответ: $C=\pi$.

Пример 2. Случайная величина X задана плотностью вероятности (распределение Лапласа) $f(x)=\frac{1}{2}\cdot e^{-|x|}$. Найти математическое ожидание величины X.

Исходя из определения математического ожидания, получим:

$M(x)=\int\limits_{-\infty}^{\infty} x\cdot f(x)dx=\frac{2}{2}\int\limits_{-\infty}^{\infty} x\cdot e^{-|x|}dx=\int\limits_{0}^{\infty} xe^{-x}dx$, исходя из четности функции $e^{-|x|}$.

Возьмём по частям неопределённый интеграл:

$$I=\int xe^{-x}dx\left\|\begin{matrix} u=x; dv=e^{-x}dx; \\ du=dx; v=-e^{-x}\end{matrix}\right\|=-x\cdot e^{-x}+\int e^{-x}dx=-x\cdot e^{-x}-e^{-x}=-e^{-x}(x+1);$$

$$M(x)=\lim_{A\to\infty}\left[-\int\limits_{0}^{A} xe^{-x}dx\right]=-\lim_{A\to\infty}\left(e^{-x}(x+1)\right)\Big|_{0}^{A}=$$

$$=-\lim_{A\to\infty}\left(e^{-A}(A+1)-1\right)=-\lim_{A\to\infty}\frac{A+1}{e^{A}}+1=1,$$

т.к. $\lim\limits_{A\to\infty}\frac{A+1}{e^{A}}=\lim\limits_{A\to\infty}\frac{1}{e^{A}}=0$ (по правилу Лопиталя).

Ответ. $M(x)=1$.

Плотность вероятности вида $\varphi(x)=\frac{1}{\sigma\sqrt{2\pi}\cdot x}\cdot e^{-\frac{(\ln x-\ln a)^2}{2\sigma^2}}$ описывает логарифмически-нормальное распределение (сокращенно логнормальное распределение) случайной величины X, т.е. её логарифм подчинён нормальному распределению. Логнормальное распределение используется для описания распределения доходов, банковских вкладов, месячной заработной платы, посевных площадей под разные культуры, долговечности изделий в режиме износа и старения.

Можно показать (замена переменной $z=\frac{\ln x-\ln a}{\sigma\sqrt{2}}$ сводит к интегралу Пуассона см. глава 12 пример 5), что математическое ожидание и дисперсия случайной величины распределённой по логнормальному закону имеют вид:

$$M(x)=\int\limits_{0}^{\infty} x\varphi(x)dx=\frac{1}{\sigma\sqrt{2\pi}}\int\limits_{0}^{\infty} e^{-\frac{(\ln x-\ln a)^2}{2\sigma^2}}dx=ae^{\frac{\sigma^2}{2}};$$

$$D(x)=\int_0^{\infty}x^2\varphi(x)dx-\left(M(x)\right)^2=\frac{1}{\sigma\sqrt{2\pi}}\int_0^{\infty}xe^{-\frac{(\ln x-\ln a)^2}{2\sigma^2}}dx-a^2e^{\sigma^2}=a^2e^{\sigma^2}(e^{\sigma^2}-1).$$

Пример 3. Проведенное исследование показало, что вклады населения в данном банке могут быть описаны случайной величиной *X*, распределённой по логнормальному закону с параметрами $a=530$, $\sigma^2=0{,}64$.

Найти средний размер вклада.

Решение. Средний размер вклада равен математическому ожиданию, т.е.

$$M(x)=\frac{1}{\sigma\sqrt{2\pi}}\int_0^{\infty}e^{-\frac{(\ln x-\ln a)^2}{2\sigma^2}}dx=ae^{\frac{\sigma^2}{2}}=530\cdot e^{0{,}64/2}=730\,(\text{ден.ед.})$$

10.6. Приложения дифференциальных уравнений к экономическим задачам

Пример 1. Скорость обесценивания автомобиля в зависимости от его возраста пропорциональна его фактической стоимости. Какова его стоимость в момент времени *t*, если его начальная стоимость S_0.

Решение. Скорость обесценивания $\frac{dS}{dt}$ – пропорциональна фактической стоимости в данный момент (знак минус, так как функция убывающая), т.е. $\frac{dS}{dt}=-kS\Rightarrow\frac{dS}{S}=-kdt$ – дифференциальное уравнение 1-го порядка с разделяющимися переменными, интегрируя которое, получим:

$$\int\frac{dS}{S}=-k\int dt\Rightarrow\ln S=-kt+\ln C\Rightarrow S=C\cdot e^{-kt}.$$

Учитывая начальные условия, получим:

$$S\Big|_{t_0}=S_0=C\cdot e^0\Rightarrow C=S_0\Rightarrow S=S_0e^{-kt}.$$

Полученное частное решение представляет экспоненциальную кривую, т.е. стоимость автомобиля убывает в зависимости от возраста по экспоненте.

Пример 2. Пусть спрос и предложение на товар определяются соответственно соотношениями

$$q_c=20p'+5p-4,\quad q_n=5p'-5p+46.$$

где p – цена товара; p' – тенденция формирования цены (производная цены по времени). Пусть также в начальный момент времени цена p за единицу товара составляла 3 ден.ед. Исходя из требования соответствия спроса предложению, найти закон изменения цены в зависимости от времени.

Для того, чтобы спрос соответствовал предложению, необходимо выполнение равенства $20p' + 5p - 4 = 5p' - 5p + 46$. Отсюда: $3p' + 2p - 10 = 0$.

Имеем дифференциальное уравнение с разделяющимися переменными:

$$3\frac{dp}{dt} = 10 - 2p,\quad \int\frac{dp}{p-5} = -\frac{2}{3}\int dt,\quad \ln|p-5| = -\frac{2t}{3} + \ln|C|,\quad p = Ce^{-\frac{2}{3}t} + 5.$$

Учтем, что $p|_{t=0} = 3$, тогда: *3=С+5; С=-2;* $p = -2e^{-\frac{2}{3}t} + 5$.

Значит, чтобы между спросом и предложением сохранилось равновесие, необходимо, чтобы цена изменялась в соответствии с полученной формулой.

Пример 3. Пусть спрос и предложение на товар определяются соотноше-ниями $q_c = 5p'' + 2p' + 5p - 2,\quad q_n = 3p'' - 2p' - 5p + 8$, где p – цена на товар; p' – тенденция формирования цены; p'' – темп изменения цены. Пусть также в начальный момент времени $p(0)=p_0$, $q_c(0)=q_c(0)=S_0$. Исходя из требований соответствия спроса предложению, найти зависимость цены от времени.

Исходя из требования соответствия спроса предложению, имеем $q_c=q_c$.

Следовательно, $5p'' + 2p' + 5p - 2 = 3p'' - 2p' - 5p + 8$, откуда получаем линейное неоднородное дифференциальное уравнение второго порядка с постоянными коэффициентами: $p'' + 2p' + 5p = 5$. Соответствующее однородное уравнение: $p'' + 2p' + 5p = 0$.

Характеристическое уравнение: $k^2 + 2k + 5 = 0$. Корни характеристического уравнения: $k_{1,2} = -1 \pm 2i$. Общее решение однородного уравнения:

$$p^*(t) = C_1e^{-t}\cos 2t + C_2e^{-t}\sin 2t.$$

Частное решение неоднородного уравнения будем искать в виде $p_{ч.н.}=A$.

Тогда $p'_{\div í\,.} = 0;\quad p''_{\div í\,.} = 0$. Подставив эти значения в дифференциальное уравнение, получим *5*А*=5,* $p_{ч.н.}=1$.

Общее решение будет таким: $p(t) = e^{-t}(C_1\cos 2t + C_2\sin 2t) + 1$.

Учтем начальные условия: $p(0)=p_0$, $q_c(0)=q_S(0)=S_0$. Тогда: $p_0=c_1+1$

$$p(t)=e^{-t}((p_0-1)\cos 2t+C_2\sin 2t)+1;$$

$$p'(t)=-e^{-t}((p_0-1)\cos 2t+C_2\sin 2t)+e^{-t}(-2(p_0-1)\sin 2t+2C_2\cos 2t)=$$

$$e^{-t}\left[(1+2C_2-p_0)\cos 2t+(-C_2-2p_0+2)\sin 2t\right];\quad p'(0)=1+2C_2-p_0;$$

$$p''(t)=-e^{-t}\left[(1+2C_2-p_0)\cos 2t-(-C_2-2p_0+2)\sin 2t\right]+$$
$$+e^{-t}\left[(2p_0-4C_2-2)\sin 2t+(-2C_2-4p_0+4)\cos 2t\right]=$$
$$=e^{-t}\left[(3-3p_0-4C_2)\cos 2t+(4p_0-3C_2-4)\sin 2t\right]\ ;\ p''(0)=3-3p_0-4C_2.$$

$$5p''(0)+2p'(0)+5p(0)-2=S_0.$$

$$15-15\cdot p_0-20\cdot C_2+2+4C_2-2p_0+5p_0-2=S_0;$$

Отсюда находим $C_2=(1/16)(S_0+12p_0-15)$.

Значит: $p(t)=e^{-t}((p_0-1)\cos 2t-\frac{1}{16}(S_0+12p_0-15)\sin 2t)+1;$

Пусть $p_0=1$, $S_0=19 \Rightarrow C_2=-1$, $C_1=0$, $p(t)=1-e^{-t}\sin 2t$.

Пример 4. Скорость распространения рекламы среди потенциальных покупателей продукции пропорциональна, как числу знающих о товаре покупателей, так и числу покупателей о нем еще не знающих. Найти закон изменения числа покупателей в зависимости от времени.

Известно, что в начальный момент времени t=0 о товаре узнало N/m человек (время отсчитывается после рекламных объявлений), m– заданное число. Найти закон изменения в зависимости от времени числа x покупателей, знающих о продукции.

Согласно условию, уравнение для определения $x=x(t)$ имеет вид $\frac{dx}{dt}=kx(N-x)$, где $\frac{dx}{dt}$ – скорость изменения числа знающих о товаре покупателей; x – число знающих о товаре; N-x – число не знающих о товаре в момент времени t; k – положительный коэффициент пропорциональности.

Начальное условие: $x\big|_{t=0}=N/m.$ Решаем дифференциальное уравнение, являющееся уравнением с разделяющимися переменными: $\frac{dx}{x(N-x)}=k\,dt.$ В результате интегрирования имеем $\frac{1}{N}\ln\left|\frac{x}{N-x}\right|=kt+C.$

Полагая $NC=C_1$, приходим к равенству $\frac{x}{N-x}=Ae^{Nkt}$, где $A=e^{C_1}$. Решим последнее уравнение относительно x: $x=N\frac{A\cdot e^{Nkt}}{A\cdot e^{Nkt}+1}=\frac{N}{1+p\cdot e^{-Nkt}}$, где $p=1/A$. Полученное уравнение называется уравнением логистической кривой.

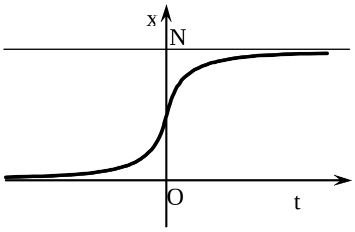


Рис. 10.6

Учтем начальные условия: $\frac{N}{m}=\frac{N}{1+p}$; $1+p=m$; $p=m-1$. Тогда $x=\frac{N}{1+(m-1)e^{-Nkt}}$ – закон изменения числа покупателей в зависимости от времени t. В частности, при m=2 получим $x=\frac{N}{1+e^{-Nkt}}$. На рисунке изображена логистическая кривая при m=2, т.е. для случая, когда ровно половина потенциальных покупателей узнала о товаре.

Пример 5.

Составить дифференциальное уравнение расширенного воспроизводства.

Обозначим: P – стоимость валового национального продукта, P_1 – стоимость средств производства, P_2 – стоимость средств потребления, S – доля перенесённой стоимости в национальном доходе. Пусть $\frac{P_1}{P}=\alpha$; $\frac{P_2}{P}=1-\alpha$, тогда $P_1=\alpha P$, $P_2=(1-\alpha)P$.

Пусть S - доля перенесённой стоимости в национальном доходе, тогда национальный доход (в стоимостном выражении) равен $P-SP=P(1-S)$. Часть национального дохода идет на увеличение производственных фондов C (в фонд накопления) с целью расширения производства. Эта часть образует скорость увеличения $\frac{dC}{dt}$ (t – время). Другая часть идет на потребление, т.е. $P(1-S)=\frac{dC}{dt}+P_2=\frac{dC}{dt}+(1-H)P$ и учитывая, что $\frac{dC}{dt}=\frac{dC}{dP}\frac{dP}{dt}=f\frac{dP}{dt}$, где $f=\frac{dC}{dP}$ – фондоёмкость прироста продукции, $P=(1-S)=f\frac{dP}{dt}+(1-\alpha)P$, откуда получим: $\frac{dP}{dt}=\frac{\alpha-S}{f}P$ – дифференциальное уравнение расширенного воспроизводства.

Задача:

Через какой промежуток времени произойдет удвоение совокупного общественного продукта P, если зависимость его от времени определяется дифференциальным уравнением расширенного воспроизводства, где $\alpha=0,7$; $S=0,4$; $f=1,5$.

При заданных постоянных параметрах уравнение принимает вид: $\frac{dP}{dt}=\frac{0,7-0,4}{1,5}P$; или $\frac{dP}{dt}=0,2P$, и является уравнением с разделяющимися переменными. Найдем общее решение: $\frac{dP}{P}=0,2dt$; $\ln P=0,2t+\ln C$; $\ln P=0,2t+\ln C$; $P=Ce^{0,2t}$. Постоянную C найдем из условия, что в начальный момент времени $t=0$ совокупный общественный продукт $P=P_0$. Очевидно, тогда $C=P_0$ и частное решение дифференциального уравнения расширенного воспроизводства $P=P_0e^{0,2\,t}$. Теперь найдем время, за которое произойдет удвоение совокупного (валового) продукта. Выразим время из частного решения $P=2P_0$, т.е.:

$$2P_0=P_0e^{0,2\,t}\Rightarrow e^{0,2t}=2\Rightarrow 0,2t=\ln 2\Rightarrow t=5\ln 2\approx 3,5.$$

Глубокая мысль дает ответ, но никто не знает, как задать вопрос таким образом, чтобы ответ мог бы быть понят.

Дуглас Адамс

РАЗДЕЛ 11. ФРАКТАЛЬНЫЙ АНАЛИЗ ФИНАНСОВЫХ РЫНКОВ

11.1. ПОНЯТИЕ ФРАКТАЛА

Одним из понятий, изменившим многие традиционные представления о геометрии, явилось понятие фрактала, введенное Бенуа Мандельбротом в 1975 году [33].

Сравнительно давно в математике возник образ объекта, более объемистого, но, тем не менее, сходного с линией. Появились геометрические формы и структуры, имеющие дробную пространственную размерность. На смену непрерывным кривым, обладающим производными произвольных порядков, пришли ломаные или очень изрезанные кривые. Так в науке возникло понятие фрактала.

Фракталами называются геометрические объекты: линии, поверхности, пространственные тела, имеющие сильно изрезанную форму и обладающие свойством самоподобия. Слово фрактал произошло от латинского слова *fractus* и переводится как дробный, ломаный. Самоподобие как основная характеристика фрактала означает, что он более или менее единообразно устроен в широком диапазоне масштабов. Так, при увеличении маленькие фрагменты фрактала получаются очень похожими на большие.

Для описания нашего физического мира древние греки создали геометрию, основанную на симметричных и гладких формах.

Многие природные системы настолько сложны, что использование только объектов классической геометрии для их моделирования представляется невозможным. Как, к примеру, построить модель горного хребта или кроны дерева в терминах геометрии?

Можно ли математически описать внезапное возникновение волны паники на финансовых рынках или даже построить математическую модель социального поведения?

Фракталы и математический хаос — подходящие средства для исследования поставленных вопросов.

Хотя универсального определения математического хаоса не существует, имеется, по-видимому, полное согласие в том, что любой вид хаоса обладает свойством непредсказуемости. Это свойство называют существенной зависимостью от начальных условий. По сути дела, математический хаос — это характерная черта именно детерминированных динамических систем. Поэтому наблюдаемые в состоянии хаоса флуктуации только кажутся случайными — их значения полностью предопределены входными параметрами. Но на практике мы никогда не располагаем абсолютно точной информацией о начальных условиях. Ошибки, пусть и ничтожные, всегда имеют место при измерении входных параметров. То, что кажется нам случайным результатом на выходе динамической системы, обусловлено большими ошибками, которые могут появиться, когда система ведет себя хаотично.

Когда-то считалось, что в детерминированной системе, при наличии достаточного объема вычислительных ресурсов, мы всегда в состояния сделать значимое предсказание (например, дать надежный прогноз погоды), несмотря на маленькие ошибки измерения текущего состояния. В присутствии хаоса это не так. Никакой самый мощный компьютер не позволит нам сделать точный прогноз на основе математической системы с существенной зависимостью от начальных условий.

Эдвард Лоренц, занимавшийся нелинейным моделированием погоды, в 1963 году обнаружил невозможность долгосрочных прогнозов погоды. Лоренц заметил, что даже ничтожные ошибки при измерении параметров текущего состояния погодных условий могут привести к абсолютно неправильным предсказаниям о состоянии погоды в будущем. Эта существенная зависимость от начальных условий лежит в основе математической теории хаоса.

11.2. ФРАКТАЛЬНОЕ ПРОСТРАНСТВО

Фрактальная геометрия в отличие от евклидовой геометрии основывается на грубости и асимметрии. Объекты не являются вариациями нескольких совершенных и симметричных форм, они бесконечно сложны. Чем более внимательно они исследуются, тем больше деталей раскрывается. Например, дерево − это фрактальная форма. Евклидова геометрия не может воспроизвести дерево. Используя евклидову геометрию, мы можем создать приближение дерева, но оно всегда выглядит искусственным, как рисунок ребенка или логотип. Евклидова геометрия воссоздает воспринимаемую симметрию дерева, но не разнообразие, которое фактически составляет его структуру. В основе такой воспринимаемой симметрии лежит хаотичность и увеличивающаяся сложность на более тонких уровнях разрешения. Качество «самоподобия» является определяющим свойством фракталов. Большинство естественных структур обладают этим свойством.

11.3. ФРАКТАЛЬНАЯ РАЗМЕРНОСТЬ

Вторая проблема, возникающая при применении евклидовой геометрии к нашему миру, — это проблема размерности. Мы живем в трехмерном мире, но только твердые формы являются действительно трехмерными, согласно определениям, которые являются основой евклидовой геометрии.

Кроме того, наше восприятие размерности может изменяться в зависимости от расстояния до объекта. С некоторого расстояния елка напоминает двумерный треугольник. Когда мы подходим ближе, она производит впечатление трехмерного конуса. Подойдя еще ближе, мы уже можем различить ее ветви, и она выглядит как сеть одномерных линий. Исследование с более близкого расстояния показывает, что ветви — это трехмерные трубки. Евклидовы структуры, напротив, становятся проще по мере увеличения масс-штаба. Трехмерное твердое тело уменьшается до двумерной плоскости. Двумерная плоскость складывается из одномерных линий и, наконец, из безразмерных точек. Наше восприятие дерева, с другой стороны, шло от двумерного к трехмерному, затем к одномерному, а затем обратно к трехмерному. Такое восприятие отличается от евклидова восприятия.

Евклидова геометрия полезна только в качестве общего упрощения мира. Фрактальная геометрия, напротив, характеризуется самоподобием и повышением сложности при увеличении. В качестве геометрии пространства она в основном применялась для создания реалистичных ландшафтов посредством компьютеров.

Следуя Бенуа Мандельброту, принимается точка зрения, согласно которой фракталы должны определяться в терминах фрактальной (дробной) размерности. Основное свойство евклидовой геометрии — это то, что размерности являются целыми числами. Прямая — это одномерный объект, а плоскость — двумерный. Тела трехмерны. Перекрутив прямую или плоскость, можно повысить размерность полученной конфигурации; при этом новая размерность обычно будет дробной. Даже гиперизмерения, развитые в более поздние времена, обладают размерностью, выражаемой в целых числах. Например, пространственно-временной континуум Эйнштейна является четырехмерным, время является четвертым измерением. Евклидовы формы гладкие, непрерывные и симметричные. Они неадекватны для описания реального мира, который они могут описывать только как общие упрощения.

11.4. ФРАКТАЛЬНАЯ МАТЕМАТИКА

Существует концептуальное различие между наукой хаоса и фракталов и евклидовой геометрией. Главное преимущество евклидовой геометрии − ее простота. С помощью евклидовой геометрии проблемы могут быть аппроксимированы и решены для нахождения оптимальных ответов. Модели формируются сравнительно простым способом, хотя они являются общими упрощениями. Фрактальная математика кажется неточной потому, что традиционные математические доказательства трудно находить и развивать. Наше понятие "доказательства' восходит к древнегреческой геометрии. Евклид создал систему аксиом, теорем и доказательства для своей геометрии. Эти понятия распространены на все остальные разделы математики. Фрактальная геометрия имеет свои доказательства, но основной метод для исследования фракталов − это метод, основанный на численных экспериментах. Используя компьютер, можно генерировать решения и исследовать фрактальные формулы. Такая "экспериментальная" форма исследования математики является новой и еще пока не заслужила уважения большинства математиков.

11.5. КЛАССИЧЕСКИЕ ФРАКТАЛЫ

11.5.1. Фрактальная размерность множества. Разделим отрезок прямой на N равных частей. Тогда каждую часть можно считать копией всего отрезка, уменьшенной в 1/r раз. Очевидно, N и r связаны соотношением Nr = 1. Если квадрат разбить на N равных квадратов (с площадью, в $1/r^2$ раз меньше площади исходного), то соотношение запишется как $Nr^2 = 1$. Если куб разбить на N равных кубов (с объемом, в $1/r^3$ раз меньше объема исходного), то соотношение примет следующий вид: $Nr^3 = 1$. Заметим, что размерность d объекта, будь то одномерный отрезок, двумерный квадрат или трехмерный куб, появляется как степень r в соотношении между N, числом равных подобъектов, и коэффициентом подобия r. А именно:

$$Nr^d = 1. \qquad (11.5.1)$$

Показатель d в равенстве (11.5.1) может не являться целым. Величину d называют фрактальной размерностью или размерностью подобия. Явное выражение для d через N и r находится логарифмированием обеих частей (11.5.1):

$$d = \frac{\ln N}{\ln 1/r}.$$

Следует иметь в виду, что понятие фрактала еще находится в развитии и разные источники могут использовать различные определения размерности.

Пусть d — обычная Евклидова размерность пространства, в котором находится фрактальный объект (d = 1 — линия, d = 2 — плоскость, d = 3 — трехмерное пространство). Покроем теперь этот объект целиком d-мерными "шарами" радиуса l. Предположим, что нам потребовалось для этого не менее чем $N(l)$ шаров. Тогда, если при достаточно малых l величина $N(l)$ меняется по степенному закону

$$N(l) = \frac{1}{l^d}, \qquad (11.5.2)$$

то d называется фрактальной размерностью этого объекта.

Формулу (11.5.2) можно переписать также в виде

$$d = -\lim_{l \to 0} \frac{\ln N(l)}{\ln l} \qquad (11.5.3)$$

Это и служит общим определением фрактальной размерности d.

Покрытие фракталов, которые обладают свойством самоподобия можно осуществлять элементами, из которых состоит данный фрактал. В этом случае имеет место упрощенный вариант формулы (11.5.3) для определения фрактальной размерности. Пусть на некотором этапе покрытия фрактала нам пришлось использовать, как минимум, $N(l)$ таких элементов характерного размера l, а на другом $N(l')$ элементов размера l'. Тогда величина фрактальной размерности d может быть вычислена по формуле:

$$d = -\frac{\ln \dfrac{N(l)}{N(l')}}{\ln\left(\dfrac{l}{l'}\right)}. \tag{11.5.4}$$

Очевидно, эту формулу можно переписать в виде $\dfrac{N(l)}{N(l')} = \left(\dfrac{l'}{l}\right)^d$, что является следствием выражения (11.5.1).

11.5.2. Пример классического фрактала

Снежинка Коха [33]. Граница снежинки, придуманной Кохом в 1904 году (рис. 11.2), описывается кривой, составленной из трех одинаковых фракталов размерности d = 1,262. Каждая треть снежинки строится итеративно, начиная с одной из сторон равностороннего треугольника. Пусгь K_0 — начальный отрезок. Уберем среднюю треть и добавим два новых отрезка такой же длины, как показано на рис. 11.1. Назовем полученное множество K_1. Повторим данную процедуру многократно, на каждом шаге заменяя среднюю треть двумя новыми отрезками. Обозначим через K_n фигуру, получившуюся после n - го шага.

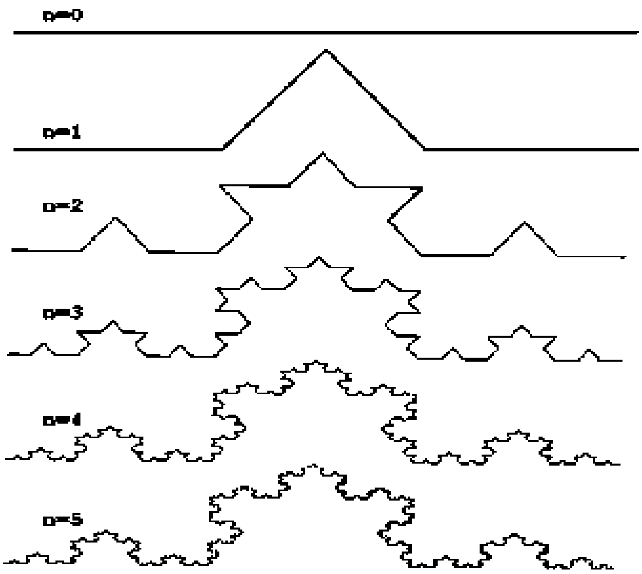


Рис.11.1 Построение кривой Коха (фрактальная размерность 1,262)

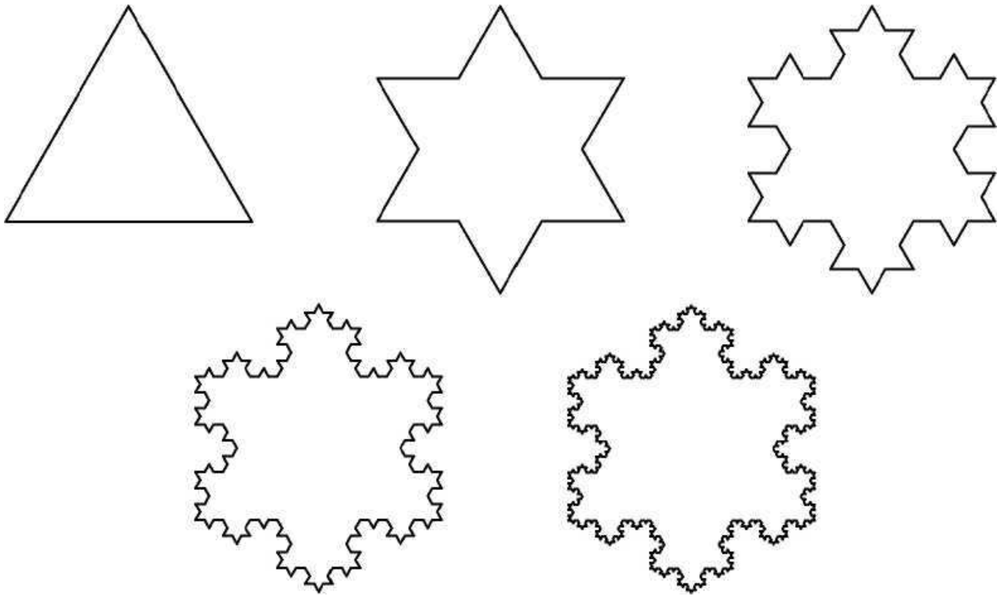


Рис 11.2 Построение снежинки Коха (фрактальная размерность 1,262)

Теорема. Граница снежинки Коха имеет бесконечную длину.

Доказательство. Достаточно показать, что каждый из трех идентичных фракталов, полученных итерациями (рис. 11.2), имеет бесконечную длину. Пусть исходный отрезок K_0 имеет единичную длину. Тогда длина кривой K_1 равна 4/3. Длина кривой K_2 равна $4^2/3^2$. Продолжая таким образом, имеем, что кривая K_n после n-го шага имеет длину $4^n/3^n$. Следовательно, длина предельной кривой K равна бесконечности.

Если знание сила, то информация,
представленная здесь,должна вести
если не к силе, то, по крайней мере, к
более высоким прибылям.

Эдгар Э. Петерс

11.6.ТЕХНИЧЕСКИЙ АНАЛИЗ ФИНАНСОВЫХ РЫНКОВ [34]

Сделки с ценными бумагами приносят участникам рынка доход (или убытки) из-за того, что цены на акции и облигации постоянно меняются, и долгое время никто не задумывался, почему так происходит. Люди просто видели результат действия и не задумывались о причинно-следственном механизме, его порождающем.

Так происходило до тех пор, пока американский финансист Чарльз Доу не опубликовал ряд статей, в которых он излагал свои взгляды на функционирование финансового рынка. Доу заметил, что цены на акции подвержены циклическим колебаниям: после продолжительного роста следует продолжительное падение, потом опять рост и падение. Таким образом, Чарльз Доу впервые заметил, что можно прогнозировать дальнейшее поведение цены на акции, если известно ее направление за какой-то последний период. Впоследствии была разработана теория технического анализа финансового рынка, которая получила название теория Доу. Эта теория ведет свое начало с девяностых годов девятнадцатого века, когда Ч.Доу опубликовал свои статьи. Поведение цены акций по Ч.Доу представлено на рис. 11.3

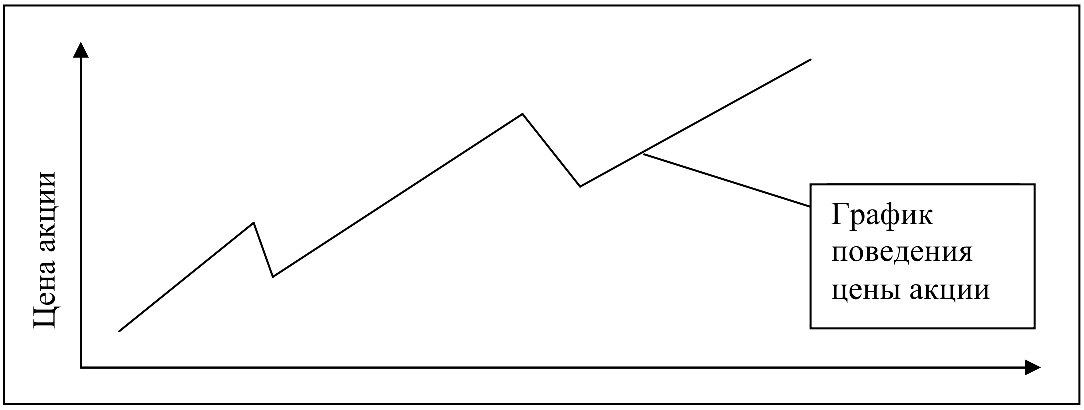


Рис 11.3 Поведение цены по Ч.Доу

Технический анализ рынков − это методы прогнозирования дальней-шего поведения тренда цены, основанные на знании предыстории его поведе-ния. Технический анализ для прогнозирования использует математические свойства трендов, а не экономические показатели ценных бумаг.

В середине двадцатого века, когда весь научный мир увлекался только что появившейся теорией фракталов, другой известный американский финансист Ральф Нельсон Эллиотт в книге «Волновой принцип» предложил свою теорию поведения цен на акции, которая была основана на использовании теории фракталов.

Основой теории служит так называемая волновая диаграмма. Следуя правилам развития массового психологического поведения, все движения цен разбиваются на пять волн в направлении более сильного тренда, и на три волны – в обратном направлении. Например, в случае доминирующего тренда мы увидим пять волн при движении цены вверх и три – при движении (коррекции) вниз.

Для обозначения пятиволнового тренда используют цифры, а для противоположного трехволнового – буквы (рис.11.4). Каждое из обратных пятиволновых движений называют импульсным, а из трехволновых − коррективным.

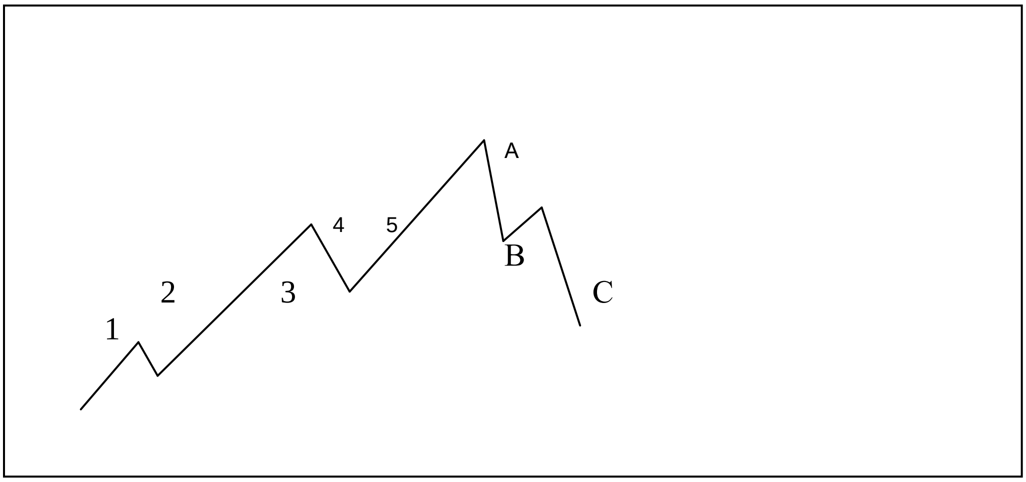


Рис 11.4

Эллиотт предположил, что каждая из только что показанных импульсных и коррективных волн также представляет собой волновую диаграмму Эллиотта. В свою очередь, те волны тоже можно разложить на составляющие и так далее. Таким образом, Эллиотт применил теорию фракталов для разложения тренда на более мелкие и понятные части. Знание этих частей в более мелком масштабе, чем самая большая волновая диаграмма, важно потому, что трейдеры, зная в какой части диаграммы они находятся, могут уверенно продавать ценные бумаги, когда начинается коррективная волна, и должны покупать их, когда начинается импульсная волна.

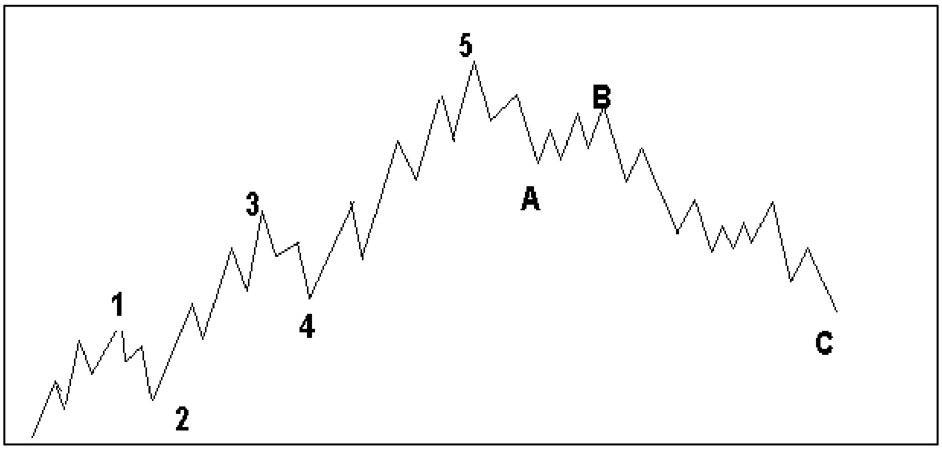


Рис.11.5. Фрактальная структура диаграммы Эллиотта

Эллиотт первым подал идею использовать числовую последовательность Фибоначчи для составления прогнозов в рамках технического анализа. С помощью чисел и коэффициентов Фибоначчи можно прогнозировать длину каждой волны и время ее завершения. Не затрагивая вопроса времени, обратимся к наиболее часто применяемым правилам определения длины Эллиоттовских волн. Под длиной в данном случае имеется в виду ее повышение или понижение по шкале цен (рис. 11.5).

Импульсные волны. Волна 3 обычно имеет длину, составляющую 1,618 волны 1, реже – равную ей. Две из импульсных волн часто бывают равны по длине, обычно это волны 5 и 1. Это происходит, если длина волны 3 меньше, чем 1,618 длины волны 1. Часто встречается соотношение, при котором длина волны 5 равна 0,382 или 0,618 расстояния, пройденного ценой от начала волны 1 до конца волны 3.

Коррекции. Длины корректирующих волн составляют определенный коэффициент Фибоначчи от длины предшествующей импульсной волны. В соответствии с правилом чередования волны 2 и 4 должны чередоваться в процентном соотношении. Наиболее распространенным примером является следующий: волна 2 составила 61,8% волны 1, при этом волна 4 может составлять только 38,2% или 50% от волны 3.

Мандельброт также предложил использовать теорию фракталов, и мультифракталы для технического анализа финансовых рынков.

На рис.11.6 представлен предложенный им генератор фрактала из трех частей (трех волн), который может быть неоднократно интерполирован в каждую часть следующих диаграмм. Появляется модель, сильно напоминающая рыночные ценовые колебания.

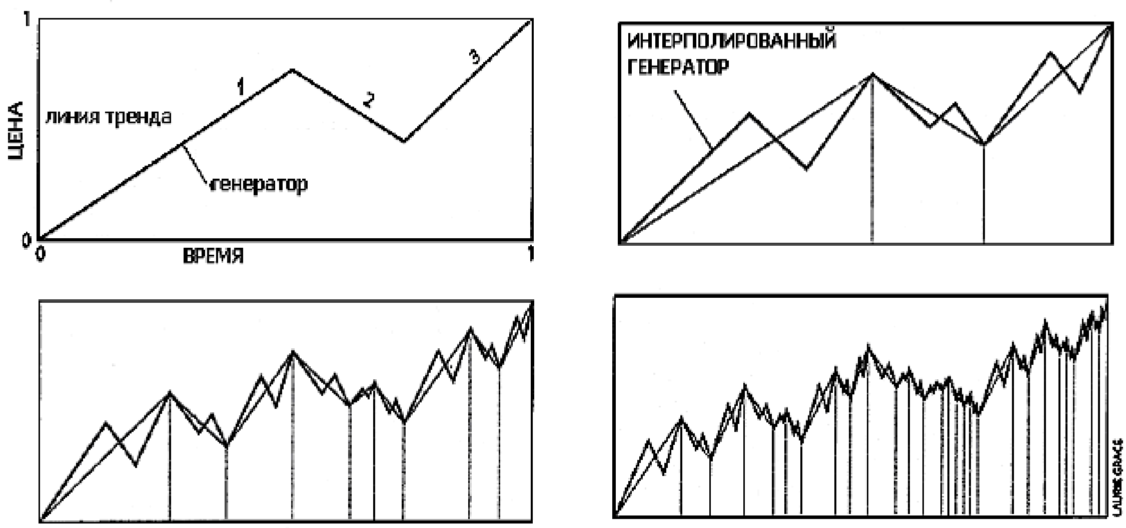


Рис. 11.6. Модель рыночных ценовых колебаний Мандельброта

Показаны только первые стадии, хотя процесс продолжает повторяться. В теории он не имеет конца, но практически бессмысленно интерполировать до интервалов времени короче, чем те, которые соответствуют интервалам между сделками, которые могут происходить по нескольку в минуту.

Отобранный генератор иллюстрирует так называемые унифрактальные кривые, которые показывают относительно спокойную картину рынка. Но спокойствие преобладает только при необычно специфических условиях. Чтобы показать обычное состояние рынка Мандельброт предложил создать мультифрактал из унифрактала. Для этого нужно удлинить или сократить горизонтальную ось времени так, чтобы части генератора были или растянуты или сжаты. В то же самое время вертикальная ценовая ось может остаться неизменной.

Эти методы не пытаются прогнозировать ценовые снижения или повышения в определенный день на основе прошлых данных, они помогают оценить вероятность того, что будет происходить на рынке.

11.7. ПРИМЕНЕНИЕ ТЕОРИИ ФРАКТАЛОВ В БУХГАЛТЕРСКОМ УЧЕТЕ

Существует множество различных компаний, форм собственности, бизнесов и организационных структур, а бухгалтерский учет – универсальный способ представления информации о финансово-хозяйственном состоянии любой организации. Разнести суммы документов по набору стандартизованных регистров учета – и состояние дел за прошедший период как на ладони. Это – с внешней точки зрения. С точки зрения, например, фискальных органов.

Для анализа происходящего внутри – все гораздо сложнее. В каждой фирме – своя организационная, своя финансовая структура, своя система ответственности и свое взаимодействие, собственные уникальные бизнес-процессы. Технология, производственные отношения, стандарты предприятия и культура производства, наконец, − составляют бесконечное разнообразие.

И под это бесконечное разнообразие должна уметь подстраиваться программа, предназначенная для учета внутренних процессов.

Этот факт, в свою очередь, объясняет разницу в сложности построения, наконец, − в стоимости между программами для автоматизации бухгалтерии и программами для автоматизации так называемого управленческого учета.

В статье о бухгалтерском учете[*] А. Блинков описал процедуру построения фрактального плана счетов учета, в котором каждый регистр не просто разбивается на линейный список субрегистров, а может содержать полную «объемную» модель бухгалтерского учета внутренних объектов (или на языке экономистов – Центров финансового учета или ответственности). Причем, процедура эта однозначно связывает исходную структуру бизнес-процессов организации с фрактальной структурой модели учета.

Есть 3 основных вида регистров: первый регистр – учет внешних связей. Хранит информацию о том, откуда или куда что-то, что имеет денежное выражение, пришло-ушло; второй регистр – хранит информацию о типе ресурса в потоке. Что именно пришло-ушло; третий регистр – для суммирования. В широком смысле суммирование есть превращение или производство. Превращение материалов в изделия, ресурсов в продукты, затрат в прибыль и т.п.

Содержащая в себе фрактальные механизмы программа универсального учета становится весьма гибкой: легко настраиваемой, декомпозируемой на рабочие места внутреннего и/или бухгалтерского учета, легко интегрируемой с любыми другими программами учета, построенными на таком же принципе.

Такая программа действительно инвариантна к предметной специфике – модель учета и модель данных она породит сама, как только пользователь "расскажет" ей о своих бизнес-процессах и своих потоках ресурсов.

11.8. СРАВНИТЕЛЬНЫЕ ХАРАКТЕРИСТИКИ ЭФФЕКТИВНЫХ И ФРАКТАЛЬНЫХ РЫНКОВ

В книге Э.Петерс [34] 2004г. гипотеза фрактального рынка излагается как альтернатива гипотезы эффективного рынка. По мнению автора фракталы присутствуют повсеместно в нашем мире и играют существенную роль, в том числе, и в структуре финансовых рынков, которые локально случайны, но глобально детерминированы. Рассмотрены методы фрактального анализа рынков акций, облигаций и валют, методы различения независимого процесса, нелинейного стохастического процесса и нелинейного детерминированного процесса и исследовано влияние этих различий на пользовательские инвестиционные стратегии и способности моделирования. Такие стратегии и

[*] Блинков А. Новое в бухгалтерском учете: фрактальность.[Электронный ресурс] –Режим доступа: http://www.xaoc.ru. 17.10.2005.

способности моделирования тесно связаны с типом активов и инвестиционным горизонтом пользователя.

В 1991 г. в книге этого же автора, озаглавленной «Хаос и порядок на рынках капитала» написано концептуальное введение к теории хаоса и фрактальной статистике. Здесь же представлены некоторые предварительные свидетельства того, что вопреки принятой теории рынки не достаточно хоро-шо описываются моделью случайных блужданий, и что гипотеза эффектив-ного рынка не достаточно хорошо подтверждается эмпирическими данными.

Гипотеза фрактального рынка задает концептуальную структуру для фрактального анализа рынка. Представление фрактальной гипотезы рынка — это основная формулировка того, как и почему функционируют рынки. Предложены также инструменты для анализа рынков в рамках фрактальной структуры.

11.8.1. Стабильность фондового рынка

Фрактальный анализ рынка − это попытка обобщить теорию рынка капитала и объяснить разнородность инвестиций. Одна из неудач традиционной теории заключается в ее попытке упростить "рынок" до среднего рационального инвестора [34].

В основе рынков лежит большая разнородность. Участие всех инвесторов не обусловлено одной и той же причиной, при этом инвесторы не используют свои стратегии на одних и тех же инвестиционных горизонтах. Стабильность рынков связана с разнородностью инвесторов. Если бы у всех участни-ков был один и тот же горизонт инвестиций, если бы они одинаково реагировали на одну и ту же информацию и вкладывали бы капитал с одной и той же целью, повсюду правила бы нестабильность. Рынки обладают стабильностью, поскольку одни инвесторы действуют ради краткосрочных прибылей, цель других − долгосрочный экономический рост. Таким образом, каждый из них диверсифицирует другого.

Фондовый рынок состоит из инвесторов, начиная с тик-трейдеров и заканчивая долгосрочными инвесторами. У каждого есть индивидуальный инвестиционный горизонт, который может быть упорядочен во времени. Стабильный рынок − это рынок, на котором все участники могут осуществлять операции друг с другом, при этом все подвергаются одинаковым рискам, независимо от инвестиционного горизонта. Показано [34], что плотность распределения прибыли одинакова для дэйтрейдеров и для недельной или даже

90-дневной прибыли, с учетом корректировки масштаба. То есть, пятиминутные трейдеры сталкиваются с тем же самым риском крупного события, что и трейдеры с недельным горизонтом.

Таким образом, рынок остается стабильным, потому что он не имеет характерного масштаба времени. Когда весь инвестиционный горизонт рынка сокращается, и каждый становится одноминутным трейдером (инвесторы потеряли веру в долгосрочную информацию), рынок становится беспорядочным и нестабильным. Следовательно, рынок может поглощать удары до тех пор, пока он сохраняет свою фрактальную структуру. Когда он теряет такую структуру, наступает нестабильность. В системах хаоса и фракталов случайность и необходимость сосуществуют. В этих системах энтропия высока, но никогда не достигает максимального состояния беспорядка из-за глобального детерминизма.

11.8.2. Несостоятельность гауссовой гипотезы

В качестве модели многомерного процесса неизвестного происхождения ученые чаще всего выбирают независимый процесс типа броуновского движения. Данная модель принимается как рабочая гипотеза. Если анализ показы-вает, что сделать прогноз трудно, гипотеза принимается как истина. Турбулент-ность жидкости моделировалась таким образом в течение многих десятилетий. Аналогично, рынки продолжают моделироваться этим же способом.

Для математиков броуновское движение имеет удобные характеристи-ки. Могут быть рассчитаны вероятности и статистические оценки с большой точностью. Однако использование традиционной статистики для моделирования рынков предполагает, что они подобны азартным играм. Каждый результат независим от предыдущих результатов. Инвестирование в ценные бумаги приравнивается к азартной игре. Но не все азартные игры обязательно управляются гауссовой статистикой. Существуют непредсказуемые системы с ограниченным числом степеней свободы. Кроме того, могут иметь место процессы, которые обладают долговременной памятью, даже если они являются вероятностными в краткосрочной перспективе.

Традиционная теория рынка капитала в значительной степени базировалась на справедливых азартных играх. Доминировала уверенность в том, что игра на бирже может быть смоделирована вероятностями. В качестве показателя риска традиционно используется такой статистический критерий как стандартное отклонение.

Гауссова статистика продолжает использоваться для моделирования цен активов.

Существовала традиция рассматривать курсы ценных бумаг и их соответственные прибыли с точки зрения спекулянта − способность инвестора получить прибыль от ценной бумаги путем предугадывания её будущей стоимости до того, как это сделают другие спекулянты. Таким образом, биржевой игрок держит пари, что текущая цена ценной бумаги выше/ниже ее будущей стоимости и, соответственно, продает/покупает ее по текущей цене. Биржевая игра предполагает заключение пари, что делает инвестиции одной из форм азартной игры. Позже Нобелевский лауреат Г. Марковиц [32] использовал стандартное отклонение как меру риска, и ковариация прибылей была использована для объяснения того, как диверсификация уменьшает риск. Приравнивание инвестиций к биржевой игре продолжилось моделью опционного ценообразования Блэка-Шоулса и другими теориями на основе равновесия. Теории биржевой игры, включая современную теорию портфеля, не проводили различия между краткосрочными биржевыми спекулянтами и долгосрочными инвесторами.

Предполагалось, что рынки "эффективны; то есть цены уже отражали всю текущую информацию, которая могла предвосхитить будущие события. Поэтому мог быть смоделирован только спекулятивный, стохастический компонент; изменение цен, вызванное изменениями стоимости, не могло быть смоделировано. Если рыночные прибыли являются нормально распределенным "белым" шумом, то они одинаковы на всех инвестиционных горизонтах. Инвестиции и азартные игры воспринимались как эквивалентные независимо от инвестиционного горизонта.

При определении статистических характеристик рынков используется нормальное распределение или известная колоколообразная кривая. Однако, известно, что рыночные прибыли не подчинены нормальному распределению.

На рис. 11.7 приведено частотное распределение 5-дневных и 90-дневных прибылей по индексу Доу-Джонса для акций промышленных предприятий со 2 января 1888г. по 31 декабря 1991г., т.е. приблизительно за 103 года. Для сравнения приведено также нормальное распределение.

Оба распределения имеют более высокие пики в среднем значении и более толстые хвосты, чем в нормальном распределении. В отличие от нормального распределения даже при четырех сигмах хвосты не сходятся к нулю.

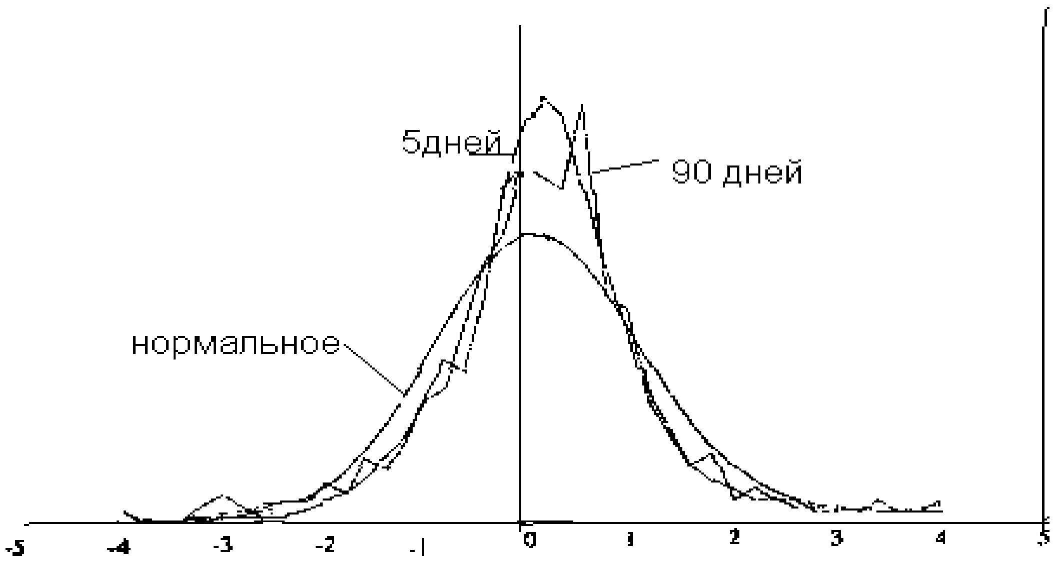


Рис 11.7 Индекс Доу-Джонса для акций промышленных компаний, частотное распределение прибылей: 1888-1991 гг.

11.8.3. Гипотеза фрактального рынка. Фрактальная статистическая структура существует, потому что она является устойчивой. Пока участвуют инвесторы с различными инвестиционными горизонтами, паника на одном горизонте может быть поглощена другими инвестиционными горизонтами в качестве возможности покупки (или продажи). Если инвестиционный горизонт становится однородным, рынок входит "в свободное падение"; появляются разрывы в последовательности ценообразования. В гауссовой окружающей среде большое изменение является суммой многочисленных небольших изменений. Во время паники рынок часто перескакивает через цены. Разрывы вызывают большие изменения, и в частотном распределении прибылей появляются толстые хвосты. Эти разрывы, в свою очередь, являются результатом недостатка ликвидности, вызванного появлением однородного инвестиционного горизонта для участников рынка. Цены отражают сочетание краткосрочной технической торговли и долгосрочной фундаментальной оценки. Таким образом, вероятно, что краткосрочные изменения цен будут более волатильными или "более шумными", чем долгосрочные. Основная тенденция на рынке отражает изменения в ожидаемом доходе на основании изменяющейся экономической среды. Краткосрочные тенденции, более вероятно, являются результатом поведения толпы. Если ценная бумага никак не связана с экономическим циклом, то не будет никакой долгосрочной тенденции. Будут доминировать торговля ценными бумагами, ликвидность и краткосрочная информация.

Гипотеза фрактального рынка говорит, что информация оценивается согласно инвестиционному горизонту инвестора. Поскольку различные инвестиционные горизонты оценивают информацию по-разному, распространение информации также будет неровным. В любой конкретный момент времени цены не могут отражать всю имеющуюся информацию, они могут отражать только ту информацию, которая важная для этого инвестиционного горизонта.

Стандартный анализ рынка предполагает, что рыночный процесс, по существу, является стохастическим. При проверке гипотезы эффективного рынка это предположение вызывает мало проблем. Большое количество исследований с использованием стандартной методологии указало на несогласованность между эффективным рынком и наблюдаемой конъюнктурой рынка; однако новые методологии также необходимы, чтобы вос-пользоваться преимуществом рыночной структуры, намеченной в гипотезе фрактального рынка.

В соответствии с гипотезой фрактального рынка рынок стабилен, если он состоит из инвесторов, охватывающих большое количество инвестиционных горизонтов. Этот факт гарантирует существование достаточной ликвидности для трейдеров.

Информационное множество больше связано с настроением рынка и техническими секторами в краткосрочной перспективе, чем в более долгосрочной перспективе. По мере увеличения инвестиционных горизонтов доминирует более долговременная фундаментальная информация. Следовательно, изменения цены могут отражать информацию, важную только для этого инвестиционного горизонта.

Цены отражают сочетание краткосрочной технической торговли и долгосрочной фундаментальной оценки. Таким образом, вероятно, что краткосрочные изменения цен будут более волатильными или "более шумными", чем долгосрочные.

В отличие от эффективного рынка гипотеза фрактального рынка утверждает, что информация оценивается согласно инвестиционному горизонту инвестора. Поскольку различные инвестиционные горизонты оценивают информацию по-разному, распространение информации также будет неровным. В любой конкретный момент времени цены не могут отражать всю имеющуюся информацию, они могут отражать только ту информацию, которая важна для определенного инвестиционного горизонта.

Фрактальный рынок основывается на той предпосылке, что он принимает различные состояния и может перемещаться между стабильным и нестабильным режимами. Гипотеза фрактального рынка объединяет эти две модели посредством использования инвестиционных горизонтов.

РАЗДЕЛ 12. ДИСКРЕТНАЯ МАТЕМАТИКА

12.1. ЛОГИКА ВЫСКАЗЫВАНИЙ

Высказывания. Логические связки. Формулы логики высказываний

Определение. Под высказыванием принято понимать языковое предположение, о котором имеет смысл говорить, что оно истинно или ложно.

Пример. “2×2=4”, “5 – простое число”, “Волга впадает в Чёрное море”. Первые два высказывания истинны, третье – ложно.

В логике высказываний интересуются не содержанием, а истинностью или ложностью высказываний (то есть их истинным значением).

Истинностные значения – истина и ложь – будем обозначать И и Л соответственно.

Определение. Множество {И, Л} называется множеством истинностных значений.

Рассмотрим логические операции над высказываниями, при которых истинностные значения составных высказываний определяются только истинностными значениями составляющих высказываний, а не их смыслом: $\neg, \wedge, \vee, \supset (\Rightarrow), \sim$ (отрицание, конъюнкция, дизъюнкция, импликация, эквивалентность).

Определение. *Отрицанием* высказывания P называется высказывание (обозначение $\neg P$), которое истинно тогда и только тогда, когда P ложно. Отрицание определяется таблицей истинности:

P	¬P
Л	И
И	Л

Определение. *Импликацией* двух высказываний P и Q называется высказывание, ложное тогда и только тогда, когда P истинно, а Q – ложно.

Импликация обозначается $\supset (\Rightarrow)$ и читается “P влечёт Q” или “если P, то Q”или “из P следует Q”

Если рассмотреть две алгебры: алгебру логики $A = (\{0,1\}, \wedge, \vee, \bar{\ }, \rightarrow, \sim)$ и алгебру логики высказываний $B = (\{Л, И\}, \wedge, \vee, \neg, \supset, \sim)$, то они будут изоморфны друг другу, так как можно поставить взаимно однозначное соответствие между множествами $0 \leftrightarrow$ Л, $1 \leftrightarrow$ И, а операции просто одинаковы.

А если алгебры изоморфны, то элементы и операции B можно переименовать так, что B совпадает с A. Из условия изоморфизма следует, что любое эквивалентное соответствие в алгебре А сохраняется в B. То есть соотношения, полученные нами в алгебре A, автоматически распро-страняются на алгебру B. В частности, изоморфизм сохраняет ассоциатив-ность, коммутативность и дистрибутивность операций.

Замечание. С логической точки зрения двоичные объекты – это вы-сказывания, которые могут быть истинными или ложными. Формулы – это составные высказывания, истинность которых определяется истинностью входящих в них элементарных высказываний (обозначаемых буквами) и логическими операциями над высказываниями.

Алфавит логики высказываний содержит следующие символы:

1) высказывательные переменные $X_1, X_2, X_3, \ldots$;

2) логические символы $\wedge, \vee, \neg, \supset, \sim$;

3) символы скобок (,) и запятую.

Определение. Слово в алфавите логики высказываний называется формулой, если оно удовлетворяет следующему определению:

1) любая высказывательная переменная −формула;

2) если A и B −формулы, то

$(\neg A), (A \wedge B), (A \vee B), (A \supset B), (A \sim B)$ – формулы;

3) только те слова называются формулами, для которых это следует из 1) и 2).

Пример. Слово $(X_1 \wedge X_2) \supset X_3 \neg X_1$ не является формулой, а слово $((X_1 \wedge X_2) \supset X_3) \neg X_1$ является.

Определение. Если любой высказывательной переменной, входящей и формулу, придать истинностные значения И и Л, то формула будет определять *истинностную функцию*, то есть функцию, определенную на множестве $\{И,Л\}$ со значением в этом множестве.

Истинностная функция может быть представлена таблицей истинности.

Пример. Формула $F = (x_1 \supset x_2) \vee (x_1 \supset (x_2 \wedge x_1))$ определяет истинностную функцию, таблица истинности которой имеет вид:

x_1	x_2	F
Л	Л	И
Л	И	И
И	Л	Л
И	И	И

12.2. РАВНОСИЛЬНОСТЬ ФОРМУЛ

Определение. Пусть A и $B-$ две формулы, зависящие от одного и того же списка переменных $< x_{i_1},...,x_{i_k} >$. Будем называть их *равносильными*, если на любой оценке списка $< x_{i_1},...,x_{i_k} >$ они принимают одинаковые значения. (То есть таблицы их истинности совпадают при одинаковом упорядочении значений переменных.)

Равносильность формул A и B будем обозначать $A \equiv B$.

Замечание. Нужно различать символы $\sim$ и $\equiv$. $\sim -$это операция, а $\equiv -$это отношение. Отношение $\equiv$ (равносильно) является отношением эквивалентности.

Основные равносильности.

1. $A \wedge B \equiv B \wedge A$ (коммутативность);

2. $A \wedge A \equiv A$ (идемпотентность);

3. $A \wedge (B \wedge C) \equiv (A \wedge B) \wedge C$ (ассоциативность)

4. $A \vee B \equiv B \vee A$ (коммутативность);

5.. $A \vee A \equiv A$ (идемпотентность);

6. $A \vee (B \vee C) \equiv (A \vee B) \vee C$ (ассоциативность);

7. $A \vee (B \wedge C) \equiv (A \vee B) \wedge (A \vee C)$ (дистрибутивность);

8. $A \wedge (B \vee C) \equiv (A \wedge B) \vee (A \wedge C)$ (дистрибутивность);

9. $A \wedge (A \vee B) \equiv A$ (I закон поглощения);

10. $A \vee (A \wedge B) \equiv A$ (II закон поглощения);

11. $\neg\neg A \equiv A$;

12. $\neg(A \wedge B) \equiv \neg A \vee \neg B$ (I закон де Моргана);

13. $\neg(A \vee B) \equiv \neg A \wedge \neg B$ (II закон де Моргана);

14. $A \equiv (A \wedge B) \vee (A \wedge \neg B)$ (I форма расщепления);

15. $A \equiv (A \vee B) \wedge (A \vee \neg B)$ (II форма расщепления).

Любая из равносильностей может быть доказана с помощью таблицы истинностей (а также из изоморфизма алгебр A и B).

Следующая группа равносильностей показывает, что одни связки могут быть выражены через другие.

16. $A \sim B \equiv (A \supset B) \wedge (B \supset A) \equiv (A \wedge B) \vee (\neg A \wedge \neg B)$;

17. $A \supset B \equiv \neg A \vee B \equiv \neg(A \wedge \neg B)$;

18. $A \vee B \equiv \neg A \supset B \equiv \neg(\neg A \wedge \neg B)$;

19. $A \wedge B \equiv \neg(A \supset \neg B) \equiv (\neg(\neg A \vee \neg B))$.

Приведем правило, с помощью которого можно переходить от одних равносильностей к другим.

Лемма 1. Пусть $A \equiv B$ и $C -$ произвольная формула, тогда
$\neg A \equiv \neg B, A \wedge C \equiv B \wedge C, C \wedge A \equiv C \wedge B, A \vee C \equiv B \vee C, C \vee A \equiv C \vee B.$
$A \supset C \equiv B \supset C, C \supset A \equiv C \supset B, A \sim C \equiv B \sim C, C \sim A \equiv C \sim B.$

Лемма 2. Пусть $A \equiv B$ и $C -$ формула, в которой выделено вхождение переменной x_i.Пусть C_A получается из C заменой этого вхождения x_i на A, а $C_B -$ из C заменой того же вхождения x_i на B. Тогда $C_A \equiv C_B$.

Утверждение 1 (правило равносильных преобразований).

Пусть $C_A -$ формула, содержащая A в качестве своей подформулы.

Пусть C_B получится из C_A заменой A в этом вхождении на B. Тогда, если $A \equiv B$, то $C_A \equiv C_B$.

Утверждение 2 (правило устранения логических символов $\supset$ и $\sim$).

Для любой формулы можно указать равносильную ей формулу, не содержащую логических символов $\supset$ и $\sim$.

Тождественно истинные формулы. Правильные рассуждения

Пусть формула A зависит от списка переменных $< x_{i_1},...,x_{i_k} >$.

Определение. Формула A называется *тавтологией* (или *тождественно истинной),* если на любых оценках списка переменных $< x_{i_1},...,x_{i_k} >$ она принимает значение И.

Определение. Формула называется *выполнимой,* если на некоторой оценке списка переменных $< x_{i_1},...,x_{i_k} >$ она принимает значение И.

Определение. Формула A называется *тождественно ложной,* если на любых оценках списка переменных $< x_{i_1},...,x_{i_k} >$ она принимает значение Л.

Определение. Формула A называется *опровержимой,* если на некоторой оценке списка переменных $< x_{i_1},...,x_{i_k} >$ она принимает значение Л.

С точки зрения логики тавтологии есть не что иное, как логические законы.

Перечислим наиболее важные тавтологии ($A, B, C -$ произвольные формулы):

1. $A \vee \neg A$ (закон исключения третьего);
2. $A \supset A$;
3. $A \supset (B \supset A)$;
4. $(A \supset B) \supset ((B \supset C) \supset (A \supset C))$;
5. $(A \supset (B \supset C)) \supset ((A \supset B) \supset (A \supset C))$;
6. $(A \wedge B) \supset A,\ (A \wedge B) \supset B$;
7. $A \supset (B \supset (A \wedge B))$;
8. $A \supset (A \vee B), B \supset (A \vee B)$;
9. $(\neg B \supset \neg A) \supset ((\neg B \supset A) \supset B)$;
10. $((A \supset B) \supset A) \supset A$ (закон Пирса).

Для доказательства 1 − 10 можно составить таблицу истинности (для разных значений A, B, C).

При доказательстве утверждений различных математических теорий обычно используют рассуждения, которые на языке логики можно выразить формулами.

12.3. ЛОГИКА И ИСЧИСЛЕНИЕ ПРЕДИКАТОВ

Логика высказываний− это узкая логическая система, в ней не все логические рассуждения могут быть осуществлены. Например, в рамках этой системы нельзя записать такое рассуждение “Простое число 2–четное. Следовательно, *существуют* простые четные числа”.

12.3.1 Предикаты, кванторы

Рассмотрим предложения, зависящие от параметров, например: “ х−четное число”, “ у меньше х”, “х−сестра у”. Поставив вместо параметров конкретные значения, мы получим высказывания, которые могут быть либо истинными, либо ложными. Например: “3– четное число”, “2 меньше 3”, “ Вера сестра Юли”. Предложения такого типа, т.е. зависящие от параметров, являются предикатами.

Определение. *Предикатом* $P\left(x_1,...,x_n\right)$ называется функция, переменные которой принимают значения из некоторого множества M, а сама она принимает два значения И (истинное) и Л (ложное), т.е. $P\left(x_1,...,x_n\right)$: $M^n \to \left\{\text{И,Л}\right\}$, где ($M^n = M\times M\times ...\times M$).

Определение. Предикат от n аргументов называется $n-$ *местным предикатом*.

Определение. Множество M значений переменных − *предметная область* предиката, x_i – свободные переменные.

Предикаты обычно обозначаются большими латинскими буквами, иногда указывают число переменных у предиката путем введения верхнего индекса. Например: $P^{(n)}\left(x_1,...,x_n\right) - n-$ местный предикат.

Над предикатами можно производить обычные логические операции.

Например: 1. Пусть $P^{(1)}\left(x\right)-$ предикат “ x делится на 3 ”, $Q^{(1)}\left(x\right)-$ предикат “x делится на 5 ”, тогда выражение $P^{(1)}\left(x\right)\wedge Q^{1}\left(x\right)$ означает “x делится на 3 и делится на 5 ”,т. е. определяет предикат делимости на 15.

2. Пусть $Q^{(2)}\left(x,y\right)$ означает предикат “$x = y$”. Он принимает значение И тогда и только тогда, когда $x = y$.

Выражение $\neg Q^{(2)}\left(x,x\right)\supset Q^{(2)}(x,y)$ определяет предикат, принимающий значение И при любых x и y.

Кроме операций логики высказываний к предикатам применяют операции связывания квантором.

Пусть $P(x)-$ предикат, где $x \in M$.

Определение. Под выражением $(\forall x)\quad P(x)$ будем подразумевать высказывание истинное, когда $P(x)$ истинно для каждого элемента x из множества M, и ложное – в противном случае.

Читается это так: “ для всех $x \quad P(x)$” или “для всех $x \in M, P(x) = И$”. Это высказывание уже не зависит от x.

Определение. Символ $\forall x$ называется *квантором общности*.

Определение. Под выражением $(\exists x)\quad P(x)$ будем понимать высказывание истинное, когда существует элемент множества M, для которого $P(x)$ истинно, и ложное −в противном случае.

Читается: “существует x такое, что $P(x)$” или “$\exists x$, для которого $P(x) = И$.

Определение. Символ $\exists x$ называется *квантором существования*.

Для предикатов в предыдущем примере:

$(\exists x)\quad (P^1(x) \wedge Q^1(x))-$ истинное высказывание.

$(\forall x)\quad (P^1(x) \wedge Q^1(x))-$ ложное высказывание ($x \in N$).

12.3.2. Формулы логики предикатов

Определим понятие *формулы* логики предикатов. Алфавит логики предикатов содержит следующие символы:

1) символы предметных переменных: $x_1,\ldots,x_n,\ldots$;

2) символы предикатов: $A_1^{(t)}, A_2^{(t)},\ldots, A_k^{(t)},\ldots$, где $t = 0,1,2,\ldots$;

3) логические символы: $\neg, \wedge, \vee, \supset, \sim$;

4) символы кванторов: $\exists, \forall$;

5) скобки и запятую:) , (.

Часто символы переменных будем обозначать через $x, y, z,\ldots,$ а символы предикатов–через P, S, Q, R и т.д.

Определение. Слово в алфавите логики предикатов называется *формулой*, если оно удовлетворяет следующему индуктивному определению.

Если $A_j^{(t)}$ – символ предиката, $x_{i1},...,x_{it}$ – символы предметных переменных (не обязательно различные), то $A_j^{(t)}(x_{i1},...,x_{it})$ – формула.

Определение. Такая формула называется *атомарной*.

Все предметные переменные формулы − свободные переменные, связанных переменных нет.

Пусть A – формула. Тогда $(\neg A)$ – формула. Свободные и связанные переменные формулы $(\neg A)$ – это свободные и связанные переменные формулы A.

Пусть A и B – формулы, причем нет таких предметных переменных, которые были бы связаны в одной формуле и свободны – в другой, тогда

$(A \vee B), (A \wedge B), (A \supset B), (A \sim B),$

есть формулы, в которых свободные переменные формул A и B остаются свободными, а связанные переменные формул A и B остаются связанными.

Пусть A – формула, содержащая свободную переменную x, тогда $(\forall x)A$, $(\exists x)A$ – тоже формулы.

Переменная x становится связанной переменной. Остальные переменные в этих формулах свободны или связаны так же, как и в формуле A.

Определение. Формула A в $(\forall x)A$ называется областью действия квантора $\forall x$.

Определение. Формула A в $(\exists x)A$ называется областью действия квантора $\exists x$.

Слово в алфавите логики предикатов 1) −5) является формулой только в том случае, если это следует из правил 1–4.

Замечание: из данного выше определения ясно, что никакая переменная не может быть одновременно свободной и связанной.

Пример 1.Следующие выражения являются формулами логики предикатов:

$A_5^{(2)}(x_1, x_7)$ – атомарная формула, x_1, x_7 – свободные переменные.

$(\forall x)(\exists x_2)A_1^{(3)}(x_1, x_2, x_3) \supset (\forall x_1)A_1^{(2)}(x_1, x_4)$ – формула, x_1, x_2 – связанные переменные, x_3, x_4 – свободные.

Выражение $(\exists x_1)(\forall x_2)A_1^{(2)}(x_1, x_3) \wedge A_2^{(2)}(x_1, x_2)$ не является формулой, т. к. x_1 в первой части выражения связанная переменная, а во второй −свободная.

Значение формулы определено лишь тогда, когда задана какая-нибудь интерпретация входящих в неё символов.

Определение. Под *интерпретацией* понимают систему $M =< M, f >$, состоящую из непустого множества M и соответствия f, сопоставляющего каждому предикатному символу $A_j^{(t)}$ определенный $t-$местный предикат (будем обозначать предикаты, поставленные в соответствие символам, теми же символами).

При заданной интерпретации считают, что предметные переменные пробегают множество M.

Для данной интерпретации каждая формула без свободных переменных представляет собой высказывание, которое истинно или ложно, а всякая формула со свободными переменными выражает некоторый предикат на множестве M, который истинен при одних значениях переменных из этого множества и ложен при других.

Определение. Определим *значение формулы в данной интерпретации,* следуя индуктивным шагам определения формулы. Значение формулы F на наборе $< a_1,...,a_n >, a_i \in M$, своих свободных переменных $x_{i_1},...,x_{i_k}$, обозначим символом $F\Big|_{(a_1,...,a_k)}$.

1. Формула $F-$ атомарная формула $A_j^{(t)}(x_{i1},...,x_{it})$. Пусть $x_{i_1},...,x_{i_k}-$ все различные свободные переменные этой формулы, записанные в определенном порядке.

Определение. *Значением формулы* F на наборе $< a_1,...,a_s >, a_j \in M$, называется значение $t-$местного предиката, сопоставленного символу $A_j^{(t)}$ при соответствующем замещении его переменных элементами $a_1,...,a_s$.

2. Формула F имеет вид $\neg A$. Пусть значение формулы A на наборе $< a_1 < a_1,...,a_n >, a_i \in M$ есть ε. Тогда $F\Big|_{<a_1,...,a_n>} = \neg\varepsilon$.

3. Формула F имеет вид $A \vee B, A \wedge B, A \supset B$ или $A \sim B$. Тогда, если $A\Big|_{<a_1,...,a_n>} = \varepsilon_1$, $B\Big|_{<a_1,...,a_n>} = \varepsilon_2$, то $F\Big|_{<a_1,...,a_n>} = \varepsilon_1 \vee \varepsilon_2$, либо $F\Big|_{<a_1,...,a_n>} = \varepsilon_1 \wedge \varepsilon_2$, либо $F\Big|_{<a_1,...,a_n>} = \varepsilon_1 \supset \varepsilon_2$, либо $F\Big|_{<a_1,...,a_n>} = \varepsilon_1 \sim \varepsilon_2$ соответственно.

4. Формула F имеет вид $(\forall x)A$. Если $x_{i1},...,x_{in}$ – совокупность всех свободных переменных формулы F, то $x, x_{i1},...,x_{in}$ – все свободные переменные формулы A.

Значение $(\forall x)A\Big|_{<a_1,...,a_n>} = И$ тогда и только тогда, когда для любого $a \in M$, $A\Big|_{<a_1,...,a_n>} = И$.

5. Формула F имеет вид $(\exists x)A$. Если $x_{i1},...,x_{in}$ – совокупность всех свободных переменных формулы F, то $x, x_{i1},...,x_{in}$ – все свободные переменные формулы A.

Значение $(\exists x)A\Big|_{<a_1,...,a_n>} = И$ тогда и только тогда, когда для некоторых $a \in M$, $A\Big|_{<a_1,...,a_n>} = И$.

Пример 1. Рассмотрим формулу $A_1^{(2)}(x_1, x_2)$. Возьмем в качестве интерпретации множество целых положительных чисел, а $A_1^{(2)}(x, y)$ интерпретируется как “$x \le y$”. Тогда формула $A_1^{(2)}(x, y)$ – это предикат “$x_1 \le x_2$”, который принимает истинное значение для всех пар a, b целых положительных чисел таких, что $a \le b$.

Пример 2. Пусть $M =< N_0, f >$, где N_0 – множество натуральных чисел и нуль $(N_0 = N \bigcup \{0\})$, f – соответствие, ставящее символу $S^{(3)}(x, y, z)$ предикат: $x + y = z$. Запишем формулы, истинные в M тогда и только тогда, когда выполнены следующие условия:

а) $x = 0$, б) x – четное.

Ответ: $F_1(x) = (\forall y)S^{(3)}(x, y, z)$, т. к. $x + y = y, \forall y$ тогда и только тогда, когда $x = 0$;

б) $F_2(x) = (\forall y)S^{(3)}(y, y, z)$, т. к. предикат $F_2(x)$ примет вид $\exists y : y + y = x$, т.е. $\exists y : 2y = x$, этот предикат принимает значение “И” тогда и только тогда, когда x является четным.

12.3.3. Равносильность формул

Пусть F и G – формулы, имеющие одно и то же множество свободных переменных (в частности пустое).

Определение. Формулы F и G *равносильны в данной интерпретации*, если на любом наборе значений свободных переменных они принимают одинаковые значения (т. е. формулы выражают один и тот же предикат).

Определение. Формулы *равносильны на множестве* M, если они равносильны во всех интерпретациях, заданных на множестве M.

Определение. Формулы F и G *равносильны* (в логике предикатов), если они равносильны на всех множествах (тогда будем писать $F \equiv G$). Для логики предикатов сохраняются равносильности и правила равносильных преобразований логики высказываний. Кроме того, можно доказать следующие правила:

1. Перенос квантора через отрицание. Пусть $A-$ формула, содержащая свободную переменную x. Тогда справедливы равносильности:

$$\neg(\forall x)A(x) \equiv (\exists x)\neg A(x);$$

$$\neg(\exists x)A(x) \equiv (\forall x)\neg A(x).$$

Доказательство. Докажем первую равносильность из первого правила. Пусть $x_{i1},\dots,x_{im}-$ множество (может быть, пустое) всех свободных переменных формулы A, причем $x_{ij} \neq x,\ j=\overline{1,m}$. Пусть $M =< M, f > -$произвольная интерпретация. Докажем, что на любом наборе значений свободных переменных $< a_1,\dots,a_m >, a_i \in M$ формулы $\neg(\forall x)A(x)$ и $(\exists x)\neg A(x)$ принимают одинаковые истинностные значения. Тогда возможны 2 случая:

1) существует элемент $a_0 \in M$, такой что $A(x)\big|_{<a_0,a_1,\dots,a_m>} =$ Л;

2) для любого элемента $a \in M$ выполняется $A(x)\big|_{<a,a_1,\dots,a_m>} =$ И.

В первом случае для элемента $a_0 \neg A(x)\big|_{<a_0,a_1,\dots,a_m>} =$ И. Поэтому $(\exists x)\neg A(x)\big|_{<a_1,\dots,a_m>} =$ И. С другой стороны, в этом случае $(\exists x)A(x)\big|_{<a_1,\dots,a_m>} =$ Л. Отсюда $\neg(\forall x)A(x)\big|_{<a_0,a_1,\dots,a_m>} =$ И. Во втором случае для любого элемента $a \in M$, $\neg A(x)\big|_{<a,a_1,\dots,a_m>} =$ Л. Отсюда можно записать, что $(\exists x)\neg A(x)\big|_{<a_1,\dots,a_m>} =$ Л. С другой стороны, в этом случае $(\forall x)A(x)\big|_{<a_1,\dots,a_m>} =$ И, откуда $\neg(\forall x)A(x)\big|_{<a_1,\dots,a_m>} =$ Л. Таким образом, первая равносильность доказа-на. Вторую равносильность можно доказать, используя первую. Для этого применим первую равносильность к формуле $\neg A(x)$. Тогда получаем $\neg(\forall x)A(x) \equiv \neg\neg(\exists x)\neg A(x) \equiv (\exists x)\neg A(x)$. Применим равносильность “снятие двойного отрицания” из основных равносильностей логики высказываний, получим $\neg(\exists x)A(x) \equiv \neg\neg(\forall x)\neg A(x) \equiv (\forall x)\neg A(x)$.

2. Вынос квантора за скобки. Пусть $A(x)-$ формула, содержащая свободную переменную $x, B-$формула, не содержащая свободную переменную x, (причем нет таких предметных переменных, которые были бы связаны в одной формуле и свободны в другой). Тогда:

$(\exists x)(A(x) \wedge B) \equiv (\exists x)A(x) \wedge B,$

$(\forall x)(A(x) \wedge B) \equiv (\forall x)A(x) \wedge B,$

$(\exists x)(A(x) \vee B) \equiv (\exists x)A(x) \vee B,$

$(\forall x)(A(x) \vee B) \equiv (\forall x)A(x) \vee B.$

Если B зависит от x, то есть только два тождества:

$(\forall x)(A(x) \wedge B(x)) \equiv (\forall x)A(x) \wedge (\forall x)B(x),$

$(\exists x)(A(x) \vee B(x)) \equiv (\exists x)A(x) \vee (\exists x)B(x).$

3. Перестановка одноименных кванторов.

$(\forall y)(\forall x)(A(x, y) \equiv (\forall x)(\forall y)A(x, y);$

$(\exists y)(\exists x)(A(x, y) \equiv (\exists x)(\exists y)A(x, y).$

(разноименные кванторы не коммутативны)

$(\forall y)(\exists x)(P(x, y) \neq (\exists x)(\forall y)P(x, y).$

4. Переименование связанных переменных. Заменяя связанную переменную формулы A другой переменной, на входящей в эту формулу, в кван-торе и всюду в области действия квантора получаем формулу, равносильную A.

Определение. Формулы, в которых из логических символов имеются только символы $\wedge, \vee$ и $\neg$, причем, $\neg$ встречается только перед символами предикатов, будем называть *приведенными формулами*.

Пример: 1) $A_1^{(1)}(x_1) \vee A_2^{(2)}(x_1, x_2)-$ приведенная формула;

2) $(\forall x_1)A_1^{(1)}(x_1) \wedge \neg A_2^{(2)}(x_2, x_3)-$ приведенная формула;

3) $\neg(A_1^{(1)}(x_1) \vee \neg A_1^{(1)}(x_2))-$ нет (отрицание перед выражением).

Теорема 1.Для любой формулы существует равносильная ей приведенная формула, причем множества свободных и связных переменных этих формул совпадают.

Определение. Приведенная формула называется *нормальной*, если она не содержит символов кванторов, или все символы кванторов стоят впереди (т.е. логические символы предикатов стоят в области действия каждого квантора).

Пример 1. $(\forall x_1)(\exists x_2)(\neg A_1^{(1)}(x_1) \vee A_1^{(2)}(x_1, x_2))-$ нормальная формула.

2. $(\forall x_1)\neg A_1^{(1)}(x_1) \wedge (\exists x_2)(A_2^{(1)}(x_2)-$ приведенная, но не нормальная формула.

Теорема 2. Для любой приведенной формулы существует равносильная ей нормальная формула той же длины (длина формулы равна сумме символов логических операций и предикатов).

Выполнимость, общезначимость. Рассмотрим некоторую интерпретацию с множеством M.

Определение. Формула A *выполнима в данной интерпретации,* если существует набор $< a_1, a_2, ..., a_n >, a_i \in M$ значений свободных переменных $x_{i_1}, ..., x_{i_n}$ формулы A такой, что $A\Big|_{<a_1,...,a_n>} =$ И.

Определение. Формула A *истинна в данной интерпретации,* если она принимает значение И на любом наборе $< a_1, a_2, ..., a_n >, a_i \in M$ значений свободных переменных $x_{i_1}, ..., x_{i_n}$.

Определение. Формула A *выполнима в логике предикатов,* если существует интерпретация, в которой A выполнима.

Определение. Формула A *общезначима* или *тождественно-истинна в логике предикатов,* если она истинна в каждой интерпретации.

Задача распознавания общезначимости формул логики предикатов сложнее, чем формул в логике высказываний (т.к. $|M| = \infty$ в общем случае). В логике предикатов (как и в логике высказываний) ставится *проблема разрешимости:* указать эффективный способ (алгоритм) распознавания общезначимости формул (т. е. является ли данная формула общезначимой или нет). В общем случае эта проблема в логике предикатов *неразрешима*.

Теорема (Черча). Не существует алгоритма, который для любой формулы логики предикатов устанавливает, общезначимая она или нет.

Замечание. В частных случаях проблема разрешимости разрешается. Например, если рассматривать формулы логики предикатов, содержащие одноместные предикатные символы, то такой алгоритм существует (такая логика существует, она описана еще Аристотелем).

Исчисление предикатов. В связи с неразрешимостью проблемы разрешимости в логике предикатов становится актуальным построение формальной теории логики предикатов – исчисления предикатов − на основании метода формальных аксиоматических теорий.

Выделение общезначимых формул (как и в исчислении высказываний) осуществляется путем указания некоторой совокупности формул, которые называются аксиомами, и указания правил вывода, позволяющих из общезначимых формул получать общезначимые.

Рассмотрим аксиоматическую теорию, которую будем называть исчислением предикатов.

1.Символы исчисления предикатов:

а) символы предметных переменных: $x_1,...,x_n,...$;

б) символы предикатов: $A_1^{(t)}, A_2^{(t)},..., A_k^{(t)},...$, где $t = 0,1,2,...$;

в) логические символы: $\neg, \supset$;

г) символы кванторов: $\exists, \forall$;

д) скобки и запятую:) , (.

2.Определение формулы то же, что и определение формулы в логике предикатов, используются только операции $\neg$ и $\supset$.

3. Аксиомы исчисления предикатов.

Каковы бы ни были формулы $A, B,$ и C, следующие формулы являются аксиомами (при этом не должно нарушаться определение формулы).

А1. $A \supset (B \supset A)$;

А2. $(A \supset (B \supset C)) \supset ((A \supset B) \supset (A \supset C))$;

А3. $(\neg B \supset \neg A) \supset ((\neg B \supset A) \supset B)$;

А4. $(\forall x_i)A(x_i) \supset A(x_j)$, где формула $A(x_i)$ не содержит переменной x_j;

А5. $A(x_i) \supset (\exists x_j)A(x_j)$, где формула $A(x_i)$ не содержит x_j;

4. Правила вывода исчисления предикатов (не должно нарушаться определение формулы):

1) правило m.p.: $\dfrac{(A, A \supset B)}{B}$;

2) правило связывания квантором общности: $\dfrac{(B \supset A(x_i))}{(B \supset (\forall x_i)A(x_i))}$, где формула B не содержит переменной x_i;

3) правило связывания квантором существования: $\dfrac{A(x_i) \supset B}{((\exists x_i)A(x_i) \supset B)}$, где формула B не содержит переменной x_i;

4) правило переименования связанной переменной.

Связанную переменную формулы A можно заменить (в кванторе и во всех вхождениях в области действия квантора) другой переменной, не являющейся свободной в *A*. Понятия вывода, теоремы, вывода из системы гипотез определяются в исчислении предикатов, как и в любой аксиоматической теории.

Теорема. (ослабленная теорема о дедукции). Если *Г*, $A \mapsto B$ и существует вывод в исчислении предикатов, построенный с применением только правила 1.(т. е. m.p.), то $\Gamma \mapsto A \supset B$.

Утверждение 1.Аксиомы исчисления предикатов – общезначимые формулы.

Утверждение 2. Формула, получающаяся из общезначимой формулы по любому из правил 1−4, является общезначимой.

СЛОВАРЬ ЭКОНОМИЧЕСКИХ ТЕРМИНОВ И ПОНЯТИЙ ФИНАНСОВОЙ МАТЕМАТИКИ

Активы банка – переданные в виде кредитов заемщикам средства различной величины и срочности.

Акции – долевые ценные бумаги, выпускаемые фирмами, компаниями, корпорациями для увеличения капитала. По ним выплачиваются дивиденды, а также они являются предметом игры на бирже. Владелец *обыкновенной акции* получает обычно процент от дохода компании, но в случае ее разорения теряет все. Это рисковая форма. Владелец *привилегированной акции* обычно получает доход, который не растет с увеличением дохода компании, но при разорении его риск потерь значительно меньше. В акциях указывается также доля участия в работе компании, способы выплат дивидендов и другие условия.

Купля и продажа акций производится через брокерские конторы, которые являются членами биржи по обмену акциями (stock exchange).

Антисипативные проценты – проценты, полученные по дисконтной или учетной ставке.

Банковский счет – ценная бумага, относящаяся к облигациям и означающая, что банк обязуется выплачивать заранее определенный процент от суммы счета.

Брутто-ставка – ставка процентов, скорректированная на инфляцию.

Валоризация – оценивание, т.е. нахождение стоимости денежных сумм в заданный момент времени.

Вексель – долговая расписка, содержащая обязательство выплатить определенную сумму (номинал векселя) в определенный срок.

Волатильность – стандартное отклонение изменений стоимости ценной бумаги.

Временной ряд – ряд наблюдений случайной величины в различные моменты времени.

Гипотеза фрактального рынка – теория, которая утверждает, что

1) рынок состоит из множества инвесторов с различными инвестиционными горизонтами и

2) информационные множества, представляющие важность для каждого инвестиционного горизонта, различны. Коль скоро рынок сохраняет такую фрактальную структуру, где отсутствует характеристическая временная шкала, он остается устойчивым. Когда горизонты рыночных инвесторов становятся одинаковыми, рынок теряет устойчивость, потому что каждый торгует исходя из одного и того же информационного множества.

Гипотеза эффективного рынка – теория, которая утверждает, что поскольку текущие цены отражают всю публичную информацию, ни один участок рынка не может иметь преимущество перед другим, тем самым извлекая сверхприбыль.

Дата эмиссии – дата, с которой начинается учет процента, начисляе-мого на ценную бумагу.

Девальвация – падение стоимости валюты относительно золота или других валют. Задача девальвации состоит в том, чтобы сделать экспортные товары более дешевыми, а импортные товары – более дорогими.

Дебитор (заемщик, должник) – лицо, получающее финансовые средства для временного их использования.

Декурсивные проценты – проценты, полученные по ставке наращения при выборе принципа расчета от настоящего к будущему.

Деноминация – номинальная стоимость ценной бумаги, т.е. та сумма денег, которая будет выплачена при ее выкупе.

Депозитные сертификаты – долговые обязательства, эмитируемые банком. Основная сумма долга в депозитном сертификате фиксируется в виде номинала, т.е. номинальной стоимости этой бумаги. В момент эмиссии депозитный сертификат продается инвестору (кредитору) по номинальной стоимости. При его погашении банк выкупает сертификат у владельца (инвестора) по цене выше номинала на сумму начисленных процентов за полный срок обращения сертификата.

Детерминизм – теория, согласно которой определенные результаты полностью предопределены.

Диверсификация – включение в портфель инвестиций ценных бумаг широкого круга компаний с целью избежания серьезных потерь в случае спада, охватившего лишь часть из секторов экономики (группирование некоррелированных или отрицательно коррелированных акций). Метод распыления средств на покупку небольших порций слабо коррелированных ценных бумаг т.е. *диверсификация* портфеля ценных бумаг, уменьшает риск и применяется там, где требуется прежде всего высокая защита капиталовложений от разорения.

Дисконт или скидка – проценты в виде разности D=S–P, где S − сумма на конец срока, P − сумма на начало срока.

Дисконтирование (сокращение) суммы – процесс уменьшения суммы денег на величину дисконта (скидки удержанных процентов) при движении во времени от будущего к настоящему.

Дисконтирование суммы S – расчет ее текущей стоимости P.

Дисконтный множитель – коэффициент, показывающий долю текущей суммы ссуды в окончательной величине долга (наращенной сумме).

Инвестиционный портфель – произвольный набор активов, включая наличные деньги.

Индекс покупательной способности денег – обратная величина индекса цен.

Индекс цен (инфляции) – коэффициент роста цен за указанный промежуток времени. Индекс цен − это отношение средневзвешенных цен определенного периода к средневзвешенным ценам базового периода (в США он рассчитывается на основе 265 товаров и услуг, которые входят в потребительскую корзину в 85 городах страны).

Инфляционная премия – корректировка ставки процентов для компенсации обесценения денег.

Капитализация процентов – присоединение начисленных процентов к сумме, служащей базой для их начисления.

Коэффициент наращения ренты – отношение наращенной суммы ренты к сумме ее годовых платежей или к размеру отдельного платежа.

Коэффициент приведения ренты – отношение современной стоимости ренты к сумме ее годовых платежей или к размеру отдельного платежа.

Кредитор (заимодавец) – лицо, представляющее в долг денежные средства или другие активы.

Математическое дисконтирование – вид дисконтирования, представляющий собой решение задачи, обратной наращению первоначальной ссуды.

Множитель наращения – коэффициент, равный отношению наращенной суммы к первоначальной.

Наращение или рост первоначальной суммы – процесс увеличения денег в связи с присоединением процентов к сумме долга.

Наращенная сумма потока платежей – сумма всех членов последовательности платежей с начисленными на них процентами к концу срока ренты.

Наращенная сумма ссуды (долга, депозита, других видов инвестированных средств) проценты – первоначальная ее сумма вместе с начисленными на нее процентами к концу срока.

Номинальная стоимость ценной бумаги – сумма денег, которая будет выплачена при ее выкупе.

Нормированная доходность – доход, который инвестор получает с каждой единицы инвестиционного капитала (т.е. с каждой гривны, доллара) в единицу времени.

Нормированная процентная ставка – стоимость единицы кредитных ресурсов в единицу времени.

Облигации – долговые обязательства, выпускаемые с целью аккумулирования капитала. По ним выплачиваются фиксированные проценты, а ссуда возвращается полностью в установленное время.

Пассивы банка – привлеченные свободные средства кредиторов в виде вкладов различной величины и срочности.

Период начисления – интервал времени, к которому относится (применяется) процентная ставка.

Портфель ценных бумаг – список ценных бумаг, находящихся в собственности физического или юридического лица.

Практика расчета простых процентов различает три варианта расчета: (1) точные проценты с точным числом дней ссуды (британская практика); (2) обыкновенные проценты с точным числом дней ссуды (французская практика); (3) обыкновенные проценты с приближенным числом дней ссуды (германская практика).

Приведение – определение любой стоимостной величины на некоторый момент времени. Если некоторая сумма приводится к более ранней дате, чем текущая, то применяется дисконтирование, если же речь идет о более поздней дате, то – наращение.

Принцип неравноценности денег – деньги, относящиеся к разным моментам времени имеют различную текущую стоимость.

Простая кредитная сделка – единовременная выдача кредита (займа, ссуды), погашаемого одним платежом в конце срока сделки и подразумевающего участие в ней двух лиц.

Процент обыкновенный или коммерческий получают, когда за базу измерения времени берут год, условно состоящий из 360 дней (12 месяцев по 30 дней в каждом).

Процент точный получают, когда за базу измерения времени берут действительное число дней в году: 365 или 366.

Процентная ставка – отношение суммы процентных денег, выплачиваемых за фиксированный отрезок времени к величине ссуды.

Процентные деньги или, кратко, проценты в финансовых расчетах – это абсолютная величина дохода от предоставления денег в долг в различных формах: помещение денег на депозитный счет, выдача ссуды, учет векселя, покупка сберегательного сертификата или облигации, продажа товара в кредит и т.д.

Проценты дискретные – начисление процентов производится дискретно, т.е. в отдельные (обычно равноотстоящие) моменты времени, причем, в качестве периодов начисления принимают год, полугодие, квартал, месяц.

Проценты непрерывные – непрерывное начисление процентов во времени.

Реинвестирование – неоднократное повторение процесса инвестирования суммы депозита вместе с начисленными на нее в предыдущем периоде процентами.

Рентабельность – отношение приведенных по ставке сравнения доходов к приведенным на ту же дату капиталовложениям.

Самоподобие – явление, состоящее в том, что малые части объекта качественно одинаковы с целостным объектом или подобны ему. В определенных детерминистических фракталах, вроде треугольника Серпинского малые его части при сравнении с целой фигурой выглядят одинаковыми. В случайных фракталах малые интервалы времени будут статистически подобны большим интервалам.

Своп, обмен − способ, позволяющий заемщику обменять один тип средств, который он может легко мобилизовать, на другой тип средств, который ему необходим, как правило, при посредничестве банка.

Сила роста δ – номинальная ставка процентов при m →∞, где m – число начислений процентов в году.

Современная величина (текущая стоимость) суммы S – величина P, найденная дисконтированием.

Современная величина потока платежей – сумма всех его членов, дисконтированных (приведенных) на некоторый момент времени, совпадающий с началом потока платежей или предшествующий ему.

Современная теория портфеля – общее название для методов количественного анализа портфеля рисковых активов, основанных на ожидаемой прибыли (или среднем ожидаемом значении) и риске (или стандартном отклонении) портфеля ценных бумаг. Инвесторы предпочитают портфель с наивысшей ожидаемой прибылью при заданном уровне риска.

Срок окупаемости – продолжительность периода, в течение которого сумма чистых доходов, дисконтированных на момент завершения инвестиций, равна сумме приведенных на этот же момент инвестиций.

Ставка номинальная – годовая ставка сложных процентов j при числе периодов начисления в году m. Тогда за каждый период проценты начисляют по ставке j/m.

Ставка процентов номинальная учетная – сложная годовая учетная ставка f, применяется при дисконтировании m раз в году. Тогда в каждом периоде, равном 1/m части года, дисконтирование осуществляется по сложной учетной ставке f/m.

Ставка процентов простая – это ставка, которая применяется к одной и той же начальной сумме на протяжении всего срока ссуды.

Ставка процентов сложная – это ставка, которая применяется к сумме с начисленными в предыдущем периоде процентами.

Ставка процентов сложная учетная – дисконтирование по сложной годовой учетной ставке осуществляется по формуле $P=S(1-d)^n$, где d −сложная годовая учетная ставка, S – дисконтируемая величина, P − современная стоимость S, n − срок дисконтирования.

Ставка учетная – ставка, применяемая для расчета процентов при учете векселей.

Ставка эффективная – годовая ставка сложных процентов, приводящая к тому же финансовому результату, что и m – разовое наращение в год по ставке j/m, где j –номинальная ставка.

Ставка эффективная учетная – сложная годовая учетная ставка, эквивалентная (по финансовым результатам) номинальной учетной ставке, применяемой при заданном числе m дисконтирований в году.

“**Тик**” – малейшее изменение ценовых колебаний на товарном рынке.

Тренд (направление, тенденция) – продолжительная тенденция изменения экономических показателей в экономическом прогнозировании.

Уравнение эквивалентности – уравнение, в котором сумма заменяемых платежей, приведенных к какому-либо одному моменту времени, приравнивается сумме платежей по новому обязательству, приведенных к той же дате. Разрабатывается при изменении условий контракта.

Учет, банковский или коммерческий учет – учет (покупка) векселей заключается в том, что банк до наступления срока платежа по векселю или др. платежному обязательству покупает его у владельца (кредитора) по цене ниже той суммы, которая должна быть выплачена по нему в конце срока, т.е. приобретает (учитывает) его с дисконтом.

Финансовой эквивалентности принцип – равенство (эквивалент-ность) финансовых обязательств сторон, участвующих в операции.

Фондовая биржа (рынок) – рынок (market), на котором продаются и покупаются ценные бумаг, а цены на них определяются спросом и предложением. Основная функция фондовой биржи заключается в том, чтобы дать возможность публичным компаниям, государству и местным органам власти привлекать капитал путем продажи ценных бумаг инвесторам.

Фондовые величины (или переменные состояния) – величины, представляющие мгновенные (относящиеся к моментам времени) значения различных финансовых характеристик, таких как стоимость, цена, курс.

Формула наращения по простым процентам или, кратко, формула простых процентов: $S = P(1+ni)$, где S – наращенная сумма, P – первоначальная сумма (ссуда), n – срок начисления процентов (срок ссуды), i – ставка процентов за единицу времени.

Фрактал – объект, в котором части некоторым образом подобны целому, т.е. отдельные составные части “самоподобны”. Примером является древовидное разветвление. В то время как каждая ветвь и каждое последовательно уменьшаемое разветвление различны, они качественно подобны структуре всего дерева.

Фрактальное размерность – число, которое количественно описывает то, как объект заполняет пространство. В евклидовой геометрии объекты сплошные и непрерывные – они не имеют отверстий или промежутков. Как таковые они имеют целочисленные размерности. Фракталы грубы и часто прерывисты, подобно скомканному куску бумаги, и поэтому имеют дробную или фрактальную размерность.

Фрактальное распределение – функция плотности вероятности, которая статистически самоподобна; это означает, что в различных интервалах времени статистические характеристики остаются одинаковыми.

Фундаментальная информация – информация об экономическом состоянии компании или экономики в целом. В рыночном анализе фундаментальная информация соотносится только с перспективой доходности фирмы.

Чистый приведенный доход – разность дисконтированных на один момент времени показателей дохода и капиталовложений.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

Основная литература

1. Финансовые расчеты в коммерческих сделках. [Текст]/ Батракова Л. Г.— М.: Финансы и статистика, 1998. — 120 с.
2. Начала финансовой математики. [Текст]/ Башарин Г. Я. — М.: Инфра-М, 1996. — 158 с.
3. Рынок ценных бумаг и производных финансовых инструментов. [Текст]/ Буренин А.Н.— М.: 1 Федеральная Книготорговая Компания, 1998—348с.
4. Финансовые вычисления для профессионалов. [Текст]/ Бухвалов А.В, Бухвалов В.В, Идельсон А.В —Спб.: Бхв-Петер-Бург, 2001—320 с.
5. Математика финансового менеджмента. [Текст]/ Ващенко Т. В.— М.: Перспектива, 1996.— 82 с.
6. Основы финансовых и коммерческих расчетов. [Текст]/ Капельян С.Я., Левкович О.Я.—Минск: НТЦ «АНИ», 1999.—224 с.
7. Методы оценки инвестиционных проектов. [Текст]/ Ковалев В.В.—М.:Финансы и статистика, 1998.—144 с.
8. Курс финансовых вычислений. [Текст]/ Ковалев В.В., Уланов В.А. —М.: Финансы и статистика, 1995.—328 с.
9. Основи фінансової математики. [Текст]/ Корж О. П. — Харків: Студцентр, 2006.—
10. Финансовая математика. Теория и практика финансово-банковских расчетов. [Текст]/ Кочович Е. — М.: Финансы и статистика, 1994.—268 с.
11. Математические методы и модели для магистрантов экономики. [Текст]/ Красс М.С., Чупрынов Б.П. — М.: Питер, 2006. —496с.
12. Математическое программирование. [Текст]/ Кузнецов В. А., Холод Н.И. — Минск: Высш шк.,1984.—
13. Математическое программирование. [Текст]/ Кузнецов Ю. Н., Кузубов В.И, Волощенко А. Б. – Минск: Высш шк.,1986.–
14. Основы финансовой и страховой математики. [Текст]/ Кутуков В.Б. – М.: Дело, 1998. –304 с.
15. Финансовая математика. [Текст]/ Малыхин В.И. – М.: Юнити, 1999. – 248 с.
16. Финансовый анализ ценных бумаг. [Текст]/ Меньшиков И.С. – М.: Финансы и статистика, 1998. – 352 с.
17. Введение в теорию фракталов. [Текст]/ Морозов А.Д – Ижевск: Ниу, 2002.– 160с.
18. Элементы теории математических моделей. [Текст]/ Мышкис А.Д. – М.: Наука, 1994. –192 с.

19. Обыкновенные дифференциальные уравнения. [Текст]/ Понтрягин Л.С. –М.: Наука, 1982.–330 с.

20. Финансовый менеджмент: теория и практика. [Текст]/ Стоянова Е.С. – М.: Перспектива, 2000. – 650 с.

21. Основы финансового анализа. Математические методы. Системный подход. [Текст]/ Родионов Я.В., Родионова С.П. – СПб.: Альфа, 1999. –590 с.

22. Техника финансово-экономических расчетов. [Текст]/ Салин В.П., Ситникова О.Ю. – М.: Финансы и статистика, 1998. –76 с.

23. Математическое моделирование. [Текст]/ Самарский А. А., Михайлов А.П. – М.: Наука, 1997. – 318 с.

24. Математический аппарат и методы формирования оптимальных параметров управления процессом функционирования строительного предприятия. [Текст]/ Торкатюк В.И., Шутенко Л.Н., Стадник Г.В., Колосов А.И., Архипова Е.С., Протопопова В.П. и др. Математический аппарат и методы Монография. – Харьков, 2007.– 827 с.

25. Фракталы. [Текст]/ Федер Е. – М.: Мир. 1991.– 254с.

26. Вексельное обращение. Российская и международная практика. [Текст]/ Фельдман А.А. – М.: Инфра-М, 1993. –304 с.

27. Депозитные и сберегательные сертификаты. [Текст]/ Фельдман А.А. – Чековое обращение. М.: Инфра-М, 1994. –168 с.

28. Государственные ценные бумаги. [Текст]/ Фельдман А. А. – М.: Инфра -М, 1994. –168 с.

29. Финансовая математика. Учебник. [Текст]/ Четыркин Е.М. 4-е изд. –М., Дело, 2004. –

30. Методы финансовых и коммерческих расчетов. [Текст]/ Четыркин Е.М. – М.: Дело Лтд, 1995. –320 с.

31. Математические методы финансового обслуживания. [Текст]/ Чуйко А.С, Шершнев В.Г. – М.: Изд-во Российской экономической академии, 1998. –124 с.

32. Теория эффективных портфелей ценных бумаг. [Текст]/ Шведов А.С. –М.: ГУ ВШЭ, 1999. – 144 с.

33. Фрактальная геометрия природы. [Текст]/ Мандельброт Б. – М.: 2002, – 666 с.

34. Фрактальный анализ финансовых рынков. [Текст]/ Эдгар Э. Петерс. – М.: 2004. –286 с.

35. Математические структуры и математическое моделирование. [Текст]/ Яглом И.М. – М.: Сов. радио, 1980. –165 с.

36. An Introduction to the Mathemathics of Finance. [Текст]/ McCutcheon J.J., Scott W.F. – Oxford: Heinemann Proffesional Publishing, 1986. – 464 p.

37. Начала финансовой математики. [Текст]/ Касимова О.Ю., Крыжановский Г.А.– М.: МНТФНИХ 1996.–120 с.

Дополнительная литература

1. Математическое программирование в примерах и задачах [Текст]/ Акулич И. Л., – Минск: Высш. шк., 1986.
2. Сборник задач по финансам и финансовому менеджменту. [Текст]/ Балабанов И.Т.– М.: Финансы и статистика, 1997. – 78 с.
3. Численные методы. [Текст]/ Бахвалов Н.С –. М.: Наука,. 1973. – 732 с.
4. Самоучитель по финансовым расчетам. [Текст]/ Бухвалов А.В., Иделъсон А.В.– М.: Мир, Пресс-сервис, 1998. –174 с.
5. Линейная алгебра и некоторые ее приложения. [Текст]/ Головина Л.И. – М.: Наука, 1975–407с.
6. Численные методы. [Текст]/ Данилина Н.И. и др.– М.: Высш шк., 1976.–368с.
7. Курс высшей математики. [Текст]/ Игнатьева А.В., Краснощекова Т.И., Смирнов В.Ф. – М : Высш. шк., 1964.
8. Сборник задач по финансовому анализу: Учеб. пособие. [Текст]/ – М.: Финансы и статистика, 1997.–128 с.
9. Начальный курс финансовой математики. [Текст]/ Медведев Г.А. – М.: ТОО «Остожье», 2000. –267 с.
10. Теоретическое и практическое пособие по финансовым вычислениям. [Текст]/ Меньшиков И.С. Мелкумов Я.С., Румянцев В. Я. – М.: Инфра-М, 1996. –336 с.
11. Курс дифференциального и интегрального исчисления. (Т.1,2,3) [Текст]/ Фихтенгольц Г.М. –М: Физматгиз, 1960. –2000с.
12. Линейная алгебра. [Текст]/ Рублев А.Н. – М : Высш. шк., 1968.

3. Интернет-ресурсы

1. Блинков А. Новое в бухгалтерском учете: фрактальность.[Электронный ресурс] –Режим доступа: http://www.xaoc.ru. 17.10.2005.

ОГЛАВЛЕНИЕ

НАУКОВЕ ВИДАННЯ

За загальною редакцією В. І. Торкатюка

Торкатюк Володимир Іванович,
Колосов Анатолій Іванович,
Бабаєв Володимир Миколайович,
Стадник Григорій Васильович,
Пан Микола Павлович,
Самойленко Микола Іванович,
Архіпова Олена Семенівна,
Протопопова Валентина Петрівна

МАТЕМАТИЧНІ МЕТОДИ І МОДЕЛІ В ЕКОНОМІЦІ

Монографія

(рос. мовою)

В авторській редакції

Комп'ютерне верстання *Н. В. Зражевська*

Комп'ютерний дизайн *Г. А. Коровкина*

Підп. до друку 17.12.2011 Формат 60 x 84 1/16
Друк на ризографі Ум.-друк. арк. 14,8
Тираж 500 пр. Зам. №

Видавець і виготовлювач:
Харківська національна академія міського господарства,
вул. Революції, 12, Харків, 61002
Електронна адреса: rectorat@ksame.kharkov.ua
Свідоцтво суб’єкта видавничої справи:
ДК №4064 від 12.05.2011